Observatoire ARGA

МЕЖДУНАРОДНЫЙ АНАЛИТИЧЕСКИЙ ДОКЛАД

Исследование №24

Deepfakes и синтетический контент: международные стандарты маркировки, происхождения и защиты пострадавших

Правовые основания, цифровые доказательства, независимая проверка и меры реагирования

Организация: Observatoire ARGA

Русская версия

21 rue de l’Aviation, 64600 Anglet, France

info@argaobservatory.org | www.argaobservatory.org

23 октября 2026 года

Содержание

Резюме исследования....4
1. Предмет исследования и границы применимости....5
2. Терминология: синтетический контент, монтаж и имперсонация....6
3. Карта участников и распределение ответственности....7
4. Правовая архитектура Европейского союза....7
5. Статья 50 AI Act: машинно-читаемая маркировка....8
6. Статья 50 AI Act: раскрытие характера дипфейка....8
7. Европейские рекомендации и добровольный кодекс 2026 года....9
8. Применимость обязательств к частным лицам и организациям....10
9. Digital Services Act и обращения к платформам....10
10. Директива (ЕС) 2024/1385 и интимные дипфейки....11
11. Франция: право на изображение и уголовные составы....11
12. Великобритания: создание интимных имитаций и платформенные обязанности....12
13. Соединённые Штаты: TAKE IT DOWN Act....12
14. Правовые режимы: сравнительный итог....13
15. Персональные данные, голос и идентифицируемость человека....13
16. Данные особых категорий и сведения об уголовном преследовании....14
17. Право на частную жизнь, выражение мнений и общественный интерес....14
18. Авторское право, смежные права и права на исполнение....15
19. Техническая модель угроз: от генерации до переиздания....15
20. Проверяемое происхождение: модель утверждений и цепочка доверия....16
21. Машинно-читаемая маркировка и её пределы....16
22. Детекторы дипфейков: чего они устанавливают и чего не могут....17
23. Базовые статистические показатели независимой проверки....17
24. Аудиофорензика и имитация голоса....18
25. Видео, фотографии и монтаж подлинного материала....18
26. Ошибки происхождения: отсутствующая и ложная криптографическая подпись....19
27. Цепочка хранения цифрового доказательства....19
28. Временная шкала, устройство и внешний контекст....20
29. Работа с платформами и уведомления о незаконном контенте....20
30. Сохранение доказательств при срочном удалении....20
31. Помощь пострадавшим и недопущение вторичной травматизации....21
32. Международное сотрудничество: пределы компетенции частной организации....21
33. Роль синтетических материалов в делах о международном розыске....22
34. Санкционные совпадения, AML и ложные корпоративные заявления....22
35. Миграция, убежище и ложные визуальные доказательства....23
36. Поддельные документы и визуальные имитации официальной переписки....23
37. Стандарт языка экспертных выводов и уровни уверенности....24
38. Протокол независимой рецензии и контроль смещения....24
39. Прозрачность методики и конфиденциальность проверяемого материала....25
40. Управление ошибками, опровержениями и правом на возражение....25
41. Профилактика в работе редакций и публичных организаций....26
42. Комплаенс разработчиков генеративных систем....26
43. Проверка устойчивости к повторному распространению....26
44. Сравнение детекторов и контрольных наборов....27
45. Риск ложного отрицания и ложного обвинения....27
46. Институциональная модель Observatoire ARGA....28
47. Этапы внедрения и критерии качества....28

48. Ограничения исследования и направления дальнейшей работы....29
49. Обманчиво подлинная запись и стратегия отрицания реальных событий....29
50. Информационная безопасность при работе с экспертными инструментами....30
51. Политическая коммуникация, выборы и публичные заявления....30
52. Недобросовестное применение маркировки и злоупотребление статусом эксперта....31
53. Экспертные ошибки при использовании визуальных эвристик....31
54. Пределы применения вероятностного экспертного вывода в суде....32
55. Параметры безопасной публичной демонстрации....32
56. Контроль качества переводов и расшифровок....33
57. Институциональные показатели и недопустимость ложной статистики....33
58. Финансирование независимого аудита и контроль качества контрактов....34
59. Реестр открытых вопросов для международной консультации....34
60. Заключение....35
Приложение 1. Матрица квалификации синтетического материала....35
Приложение 2. Разделение ролей в соответствии со статьёй 50 AI Act....36
Приложение 3. Матрица дат и действия нормативных требований....36
Приложение 4. Карточка первичного сообщения об инциденте....37
Приложение 5. Минимальный цифровой паспорт файла....38
Приложение 6. Реестр цепочки хранения....38
Приложение 7. Протокол проверки C2PA и других сведений о происхождении....39
Приложение 8. Паспорт проверки машинной маркировки....39
Приложение 9. Бланк независимого теста детектора....40
Приложение 10. Проверка статистики на условном примере....40
Приложение 11. Контрольный лист аудиозаписи....41
Приложение 12. Контрольный лист видеозаписи....41
Приложение 13. Бланк уведомления по статье 16 DSA....42
Приложение 14. Срочная инструкция по интимным материалам....42
Приложение 15. Сравнительный лист механизмов защиты в ЕС, Франции, США и Великобритании....43
Приложение 16. Паспорт правомерной обработки персональных данных....43
Приложение 17. Контроль публикации аналитического заключения....44
Приложение 18. Матрица конфликтов интересов эксперта....44
Приложение 19. Техническая карта риска для финансовых операций....45
Приложение 20. Структура доказательственного досье по международному розыску....45
Приложение 21. Протокол оценки материалов для убежища....46
Приложение 22. Контроль фальшивых официальных писем и скриншотов....47
Приложение 23. Рабочий протокол апелляции и исправления отчёта....47
Приложение 24. Условия экспертного поручения и границы исследования....48
Приложение 25. Матрица критериев запуска пилотной программы ARGA....48
Приложение 26. Дорожная карта на 90 дней....49
Приложение 27. Условный кейс: голосовое распоряжение о переводе средств....49
Приложение 28. Условный кейс: фальшивое признание по международному делу....50
Приложение 29. Условный кейс: компрометирующая интимная имитация....50
Приложение 30. Условный кейс: сфабрикованное официальное письмо....51
Приложение 31. Шкала доказательственных выводов....51
Приложение 32. Реестр рисков и корректирующих мер....52
Приложение 33. Контрольный лист научной рецензии....52
Приложение 34. Перечень запрещённых методических сокращений....53
Приложение 35. Термины для единообразного использования....53
Источники и нормативные материалы....54

Резюме исследования

Синтетический аудиовизуальный контент уже нельзя рассматривать только как разновидность технологической подделки. В одних обстоятельствах генерация изображения, голоса или видеозаписи служит очевидной художественной задачей; в других она создаёт ложное доказательство, способствует незаконному преследованию, имитирует публичное заявление или нарушает право человека на личную и интимную жизнь. Правовой результат зависит не только от применённой модели. Необходимо установить, кто создал материал, кто разместил его в конкретном контексте, какая информация сопровождала публикацию, какой ущерб возник и какие нормы применимы к участникам этой цепочки.

На 23 октября 2026 года в Европейском союзе действуют требования прозрачности статьи 50 Регламента (ЕС) 2024/1689. Они различают машинно-читаемую маркировку продукции генеративных систем, обязанность пользователя-профессионала раскрывать характер дипфейка и отдельные правила для публикаций текста по вопросам общественного интереса. Подготовленные в 2026 году рекомендации Комиссии и добровольный кодекс практики помогают применять эти обязанности, но не подменяют закон. Самостоятельные режимы защиты пострадавших развиваются через европейские и национальные уголовные нормы, правила работы интернет-платформ, защиту персональных данных и процессуальные гарантии. [01] [02] [03]

Исследование показывает, что единого технического признака, который доказал бы подлинность любого цифрового материала, не существует. Криптографически проверяемая история происхождения, встроенные маркеры, водяные знаки и алгоритмические детекторы отвечают на разные вопросы. Подпись может подтверждать целостность заявления определённого издателя, но не достоверность самого события. Отсутствие подписи не доказывает генерацию; отрицательный результат детектора не доказывает подлинность. Поэтому независимый аудит должен соединять техническое исследование с внешней проверкой источника, контекстом распространения и юридической оценкой.

Предлагаемая модель для Observatoire ARGA строится на трёх защищённых контурах. Первый предназначен для профилактики: правила маркировки, настройки публикации и проверка происхождения. Второй охватывает быстрые меры помощи пострадавшим и сохранения доказательств: идентификация URL, минимальное документирование, уведомление платформ и защита конфиденциальных материалов. Третий посвящён независимому разбору спорных сведений, включая случаи, когда синтетический материал попадает в досье о санкциях, международном розыске, миграции или финансовом мониторинге. Здесь необходима не автоматическая вера детектору, а оспоримая, воспроизводимая цепочка выводов.

Доклад содержит правовую сравнительную оценку, техническую модель независимой проверки, стандартизированные процедуры и рабочие формы. Аналитические рекомендации являются предложениями ARGA Observatory, а не утверждением о существовании у ассоциации публично-властных или судебно-экспертных полномочий. Все условные кейсы прямо обозначены как учебные. Информация о зарубежном праве приведена по состоянию на 23 октября 2026 года и требует проверки перед применением в конкретном производстве.

Основные выводы

1. Маркировка контента, установление его происхождения и установление истинности события требуют раздельных доказательств. 2. Европейские требования статьи 50 AI Act уже действуют, но имеют точный круг адресатов и установленные исключения; универсальной обязанности любого гражданина маркировать каждое изображение из них не следует. 3. Вред от синтетического контента определяется не только качеством имитации, но и контекстом: интимное изображение, фальсифицированная запись публичного лица и манипуляция доказательством требуют разных мер. 4. Реакция платформ и регуляторов должна быть быстрой, однако сама скорость не оправдывает уничтожения единственного экземпляра доказательств без защищённого сохранения необходимых сведений. 5. Независимая организация вправе публиковать методику и квалифицированные оценки, но не вправе выдавать техническое заключение за судебное признание подделки или устанавливать уголовную ответственность. 6. Для ARGA наибольшую практическую ценность имеет двухэтапная проверка: сначала ограничить потенциальный вред, затем обеспечить независимую проверку и право на возражение.

1. Предмет исследования и границы применимости

Под deepfake для целей исследования понимается визуальный или звуковой материал, искусственно созданный либо существенно изменённый с использованием ИИ так, что его можно принять за аутентичную запись человека, предмета, места или события. Это рабочее аналитическое определение шире или уже конкретных юридических дефиниций в зависимости от вопроса. Например, статья 3(60) AI Act содержит собственные признаки сходства с существующим объектом и ложного впечатления подлинности. Нельзя автоматически переносить эту дефиницию на состав преступления во Франции или в другом государстве. [01]

Объектом исследования являются четыре процесса: производство синтетического контента; его публикация и повторное распространение; выявление, проверка и доказательственное использование; предотвращение вреда и восстановление нарушенных прав. Отдельно рассматриваются гибридные материалы, где изменена лишь часть подлинной записи: переставленные реплики, сгенерированная мимика, подменённая озвучка, синтез лица на подлинном теле, монтаж реальных видеофрагментов в недостоверную хронологию. В этих ситуациях дихотомия «настоящий/ненастоящий» часто недостаточна.

Исследование не приравнивает всякое использование генеративных средств к противоправному поведению. Художественные фильмы, дублирование, доступная коммуникация, сатира и восстановление архивов могут иметь законную цель. Однако заявленный творческий характер не устраняет требований, предусмотренных законом, и не освобождает от оценки согласия, публичного контекста и возможных нарушений прав личности. Для каждого случая нужен отдельный анализ создателя, распространителя, адресата, цели и характера предполагаемого ущерба.

Прикладная рамка ARGA ограничивается независимой аналитикой, правовой координацией и методической поддержкой. Доклад не предлагает тайный мониторинг устройств, несанкционированное получение цифровых следов или сбор интимных изображений в общедоступный архив. Цель состоит в установлении условий, при которых уже законно полученные сведения могут быть использованы, проверены и представлены компетентному лицу.

2. Терминология: синтетический контент, монтаж и имперсонация

Синтетическим может быть материал, полностью созданный моделью, либо частично преобразованный исходник. Монтаж шире AI-модификации: обычная редактура видео и композитинг могут создавать такой же ложный контекст, как генеративная система. Имперсонация обозначает использование голоса, внешности или идентификатора для создания впечатления, что сообщение сделано определённым человеком. Мошенничество, подмена личности и унижение достоинства могут существовать даже без технически совершенного дипфейка.

Для отчётности целесообразно описывать материал по четырём независимым осям. Первая: тип преобразования, например генерация, замена, наложение, склейка. Вторая: идентифицируемость человека либо объекта. Третья: коммуникативная функция публикации, например шутка, реклама, официальное сообщение, доказательство. Четвёртая: правовой и фактический риск: обман, вторжение в частную жизнь, дискриминация, шантаж, общественная дезинформация, ошибочное правоприменение. Эта схема уменьшает соблазн считать технологический признак единственным правовым тестом.

Особое значение имеет контекст. Зритель может понимать, что анимационный персонаж вымышленный, но не распознать голосовую запись, которая имитирует распоряжение должностного лица. То же изображение, опубликованное с заметным предупреждением и затем распространённое после обрезки предупреждения, создаёт разный информационный эффект. Поэтому фиксировать следует не только байты файла, но и экран, подпись, заголовок, страницу, время и способ повторного размещения.

3. Карта участников и распределение ответственности

В технологической цепочке минимум пять возможных ролей. Разработчик создаёт модель или программную систему; поставщик делает инструмент доступным на рынке; пользователь или оператор формирует и публикует материал; платформа размещает, индексирует либо рекомендует его; проверяющий исследует источник и обстоятельства. Роли могут совпадать у одной организации, но их нельзя механически смешивать, поскольку применимые обязанности зависят от фактических функций.

В частности, статья 50 AI Act проводит разграничение между provider и deployer. Поставщик генеративной системы отвечает за технические возможности маркировки в установленных пределах, а профессиональный пользователь несёт самостоятельные обязанности раскрыть искусственный характер отдельных публикаций. Это не даёт основания возложить на поставщика любое нарушение конечного распространителя или, наоборот, освободить конечного распространителя при наличии машинного маркера. [01] [02]

Отдельный участник — вторичный републикатор, который может удалить маркировку, изменить подпись либо представить сатирический фрагмент как документальный. Для него существенны знание о характере материала, самостоятельная редакционная роль, умысел и применимое право. В заключении следует фиксировать не предполагаемую «единую ответственность за дипфейк», а конкретное действие конкретного субъекта и нормативное основание возможной обязанности.

4. Правовая архитектура Европейского союза

Регулирование синтетического контента в ЕС многослойно. Регламент об ИИ определяет отдельные обязанности прозрачности при генерации и использовании определённых систем. Digital Services Act регулирует процедуры обращения с предполагаемо незаконным контентом в сфере посреднических услуг. GDPR применяется при обработке персональных данных, включая идентифицируемое лицо, голосовые и иные сведения при выполнении критериев персональных данных. Авторское право, гражданская и уголовная ответственность дополняют эту конструкцию.

Ошибочно утверждать, что нарушение требований маркировки по статье 50 автоматически означает уголовную противоправность самого произведения. Равным образом прозрачное раскрытие синтетического происхождения не делает законным материал, нарушающий нормы об интимных изображениях, частной жизни или мошенничестве. Каждая группа норм регулирует собственный объект. В практике это требует матрицы квалификации, позволяющей одновременно рассматривать прозрачность, правомерность данных, платформенное реагирование и средства защиты конкретного лица.

Нормативный статус должен фиксироваться на дату анализа. Рекомендации Комиссии по статье 50 от 20 июля 2026 года помогают интерпретировать закон, а добровольный Code of

Practice, опубликованный 10 июня 2026 года, содержит практические решения. Обязанности возникают из действующего Регламента, а не из самого факта подписания добровольного кодекса. [01] [02] [03]

5. Статья 50 AI Act: машинно-читаемая маркировка

Часть 2 статьи 50 предусматривает, что поставщики систем, способных генерировать синтетические аудио-, визуальные, видео- и текстовые материалы, должны обеспечивать маркировку результата в машиночитаемом формате и обнаружимость искусственного происхождения или изменения. Закон ориентирует технические решения на эффективность, совместимость, устойчивость и надёжность в пределах технической осуществимости, с учётом вида контента, затрат и уровня технологий. Это обязательство отличается от визуальной надписи на опубликованной странице. [01]

Предусмотрены конкретные границы: вспомогательное стандартное редактирование и несущественное изменение исходных данных либо их смысла может находиться вне соответствующего требования; отдельно предусмотрены случаи законно разрешённой деятельности по выявлению, предупреждению и расследованию преступлений. Эти пределы должны оцениваться узко по фактам, а не превращаться в общее освобождение всех корпоративных инструментов монтажа.

Общий срок применения статьи 50 — 2 августа 2026 года. По консолидированным материалам Комиссии для определённых систем, размещённых на рынке до этой даты, предусмотрен ограниченный переходный период до 2 декабря 2026 года именно в отношении технической маркировки и обнаружения по статье 50(2). Он не отменяет весь набор требований для любых систем и не означает ретроактивной обязанности маркировать каждую старую публикацию. [02] [04]

На практике проверяющий должен запросить у поставщика точную версию модели, способ вывода отметок, условия их сохранения при экспорте, сведения об известных способах удаления и результаты испытаний. Утверждение «система соответствует статье 50» без таких сведений является заявлением, а не независимым доказательством соответствия.

6. Статья 50 AI Act: раскрытие характера дипфейка

Часть 4 статьи 50 возлагает на профессионального оператора, который использует систему ИИ для генерации либо манипулирования изображением, аудио или видео, образующими deepfake, обязанность раскрывать искусственный характер материала. Раскрытие должно быть ясным, различимым и доступным для заинтересованных людей не позднее первого взаимодействия или предъявления содержания. Оно не сводится к невидимому цифровому водяному знаку. [01] [02]

Для очевидно художественных, творческих, сатирических, вымышленных или аналогичных произведений закон допускает особый способ раскрытия, не препятствующий просмотру либо восприятию произведения. Это исключение касается формы раскрытия, а не универсального снятия защиты частной жизни, согласия или иных требований права. Отдельно предусмотрены случаи законно санкционированной правоохранительной деятельности. Для материалов, создающих риск ложного приписывания слов реальному лицу, надёжный подход — заметная отметка непосредственно возле контента, а не только в глубине пользовательского соглашения.

В отношении текстовых публикаций, предназначенных информировать общественность по вопросам публичного интереса, действуют специальные правила раскрытия с учётом правовой редакционной ответственности и человеческого контроля. Нельзя описывать их как всеобщую обязанность маркировать каждый текст, подготовленный с помощью ИИ. Внутренние правила организации при этом могут добровольно устанавливать более высокую прозрачность без утверждения, будто они прямо предписаны статьёй 50 каждому участнику.

7. Европейские рекомендации и добровольный кодекс 2026 года

Комиссия опубликовала рекомендации о сфере применения статьи 50 20 июля 2026 года, а окончательный добровольный кодекс маркировки и раскрытия синтетического контента — 10 июня 2026 года. Эти документы формируют важный технический и организационный ориентир. Кодекс имеет отдельные разделы для поставщиков, отвечающих за машинно-читаемую маркировку, и операторов, отвечающих за раскрытие дипфейков и определённых публичных текстов. [02] [03]

При оценке добросовестности организации проверяющий может сопоставлять фактическую систему маркировки с рекомендованными практиками, но обязан различать два вывода. Первый: техническое решение реализует определённые положения кодекса. Второй: выполнены юридические обязанности конкретного адресата в контексте конкретной публикации. Между ними нет автоматического равенства. Также нельзя утверждать, будто присоединение к кодексу заменяет контроль компетентного органа или создаёт иммунитет от жалоб.

Для международного Observatoire особую ценность представляют открытость методики, указание версии проверяемого программного инструмента, контроль доступности маркировки для людей с нарушением зрения или слуха и оценка того, сохраняется ли информация после публикации в различных каналах. Отчёт должен отдельно выявлять «технически нанесено», «увидит средний адресат» и «сохранится при обычном повторном использовании».

8. Применимость обязательств к частным лицам и организациям

Нельзя механически переносить AI Act на каждое действие человека с личным телефоном. Материалы Комиссии разграничивают использование в личных непрофессиональных целях и использование при хозяйственной либо профессиональной деятельности. Профессиональный статус определяется реальными функциями и обстоятельствами, а не только наличием официальной регистрации. Корпоративный создатель рекламной кампании и сотрудник, действующий в пределах полномочий, могут нести разные роли, как и независимый подрядчик. [02]

Это разграничение не означает, что личное создание и распространение интимной имитации законно. Применяться могут национальные уголовные нормы, гражданские средства защиты, правила платформ и нормы о персональных данных в зависимости от фактов. Поэтому первая юридическая проверка должна разделять вопрос применимости статьи 50 и самостоятельный вопрос противоправности вредного материала.

Внутренний комплаенс-процесс полезно организовать так: определить организационного пользователя; установить цель; зафиксировать публикационный канал; проверить наличие идентифицируемого реального объекта; оценить риск ложного впечатления подлинности; проверить наличие предупреждения для людей и технической отметки; проверить специальное отраслевое право. Решение не должно состоять только из формальной галочки «AI used».

9. Digital Services Act и обращения к платформам

Статья 16 Регламента (ЕС) 2022/2065 устанавливает механизм уведомлений о предполагаемо незаконном контенте для поставщиков хостинга. Сообщение должно быть достаточно точным и обоснованным, включать объяснение незаконности, конкретное электронное местонахождение и необходимые сведения заявителя, с предусмотренными законом исключениями. При получении надлежащего уведомления платформа должна рассмотреть его своевременно, объективно и добросовестно, а также сообщить о решении и доступных способах обжалования. [05]

DSA не устанавливает общее правило, по которому любая помеченная «deepfake» запись должна быть удалена в течение 24 либо 48 часов. Реакция зависит от характера незаконности, типа сервиса, применимого права, специальных режимов и содержания уведомления. Ложное или недостаточно конкретное заявление может затруднить удаление и затронуть законную речь. Поэтому для обращения следует указывать, какая именно правовая норма предположительно нарушена и что в материале создаёт соответствующий риск.

Проверяющий должен сохранять разделение между фактом подачи жалобы, подтверждением её получения, решением платформы, фактической недоступностью URL и прекращением дальнейшего распространения. Удаление исходного поста не равнозначно удалению копий и

не доказывает наступление юридической ответственности автора. Статус trusted flagger по статье 22 DSA также является специальным признанным статусом; не следует приписывать его ARGA только на основании экспертной деятельности.

10. Директива (ЕС) 2024/1385 и интимные дипфейки

Директива (ЕС) 2024/1385 обязывает государства-члены установить уголовно-правовые меры в отношении определённых форм насилия, включая распространение без согласия интимных материалов и отдельные манипулированные изображения, создающие впечатление сексуального действия. Статья 5 использует специальные элементы деяния, включая создание или изменение материала, последующее предоставление общедоступного доступа и порог серьёзного вреда в предусмотренных случаях. Не всякое редактирование изображения приравнивается к указанному составу. [06]

По официальной информации Комиссии срок транспонирования Директивы установлен на 14 июня 2027 года. Следовательно, по состоянию на октябрь 2026 года её следует анализировать как действующий акт ЕС, адресующий обязанности государствам по внедрению соответствующих мер, а не как единый уже полностью действующий национальный уголовный состав во всех государствах-членах. Национальное право отдельных стран может предоставлять защиту раньше и шире.

Статья 23 предусматривает меры для быстрого удаления или ограничения доступа к определённым незаконным материалам, включая возможность обязательных предписаний компетентных органов. Практическая реализация требует национальных процедур. Поэтому в сравнительных таблицах необходимо отдельно указывать обязательство государства, доступное заявителю текущее средство, компетентный орган и возможную дату применения.

11. Франция: право на изображение и уголовные составы

Во Франции статья 226-8 Уголовного кодекса охватывает распространение без согласия определённого монтажа с изображением или словами человека, когда монтаж не является явно узнаваемым или его характер не обозначен. После изменений 2024 года к предусмотренному законом поведению приравнивается распространение нераскрытого алгоритмически созданного визуального или звукового содержания, представляющего изображение или слова лица. Норма содержит конкретные условия и повышенные санкции за распространение через онлайн-сервис. [07]

Отдельная статья 226-8-1 касается сексуализированных монтажей и алгоритмически созданного визуального или звукового материала сексуального характера с образом либо словами человека без согласия. Для этой статьи предусмотрены иные санкции; на стадии квалификации её нельзя просто заменить общей нормой статьи 226-8. Взаимоотношение уголовного и гражданского способов защиты следует рассматривать по обстоятельствам: изображение, частная жизнь, репутация, вред и средства прекращения распространения. [08]

Практический вывод для ARGA: в жалобе не следует ограничиваться общей формулой «нарушен AI Act». Следует установить наличие согласия, публикации третьим лицам, восприятие материала аудиторией, характер изображения и обстоятельства размещения. Решение о наличии уголовного состава принимает компетентный орган; аналитическая организация может сформировать фактическую хронологию и указать возможные юридические основания, избегая категоричного вывода о виновности неизвестного лица.

12. Великобритания: создание интимных имитаций и платформенные обязанности

В Великобритании правила защиты от интимного насилия претерпели существенные изменения к 2026 году. По информации Crown Prosecution Service, с 6 февраля 2026 года действуют новые положения Sexual Offences Act 2003, относящиеся к созданию или запросу создания определённых мнимых интимных изображений взрослого без согласия либо разумной уверенности в согласии. С 29 июня 2026 года отдельные положения распространились на производство либо предоставление инструментов создания подобных изображений. [09]

Эта уголовно-правовая рамка не совпадает с европейским регулированием маркировки. Преступность может быть связана с созданием или целевым запросом материала, даже если он ещё не разошёлся по большой аудитории; при этом конкретные элементы нормы, лицо и дата действия требуют проверки. Правила не следует описывать так, словно каждое несогласованное использование фотографии в любительском коллаже автоматически образует указанный состав.

В сфере онлайн-безопасности Ofcom объявлял усиленные меры против незаконных интимных изображений и рекомендовал автоматическое сопоставление цифровых отпечатков либо равноэффективные меры при соблюдении соответствующих кодексов. При оценке комплаенса нужно отличать правовую обязанность конкретного сервиса, рекомендации кодекса, дату вступления соответствующей меры в силу и технический метод исполнения. Отсутствие определённого коммерческого продукта для сравнения хешей не доказывает автоматически нарушение, если допустима иная равноэффективная система. [10]

13. Соединённые Штаты: TAKE IT DOWN Act

Федеральный закон США TAKE IT DOWN Act был принят в мае 2025 года. С 19 мая 2026 года требования раздела 3 обязывают определённые онлайн-платформы предоставить механизм обращения и по надлежащему уведомлению удалить охватываемые законом интимные изображения и известные идентичные копии в течение 48 часов. Закон распространяется на предусмотренные им «digital forgeries» наряду с реальными изображениями. Федеральная торговая комиссия заявляет о надзорной функции в части соответствующих обязательств платформ. [11]

Значение этого правила ограничено предметом, адресатом и условиями закона. 48-часовой срок не относится к любому политическому видео, сатире, синтетическому голосу или публикации в любой юрисдикции. Для использования механизма необходимо установить, является ли сервис covered platform, отвечает ли материал определению, допустимо ли обращение данного заявителя и получено ли действительное уведомление. Следует отдельно учитывать уголовно-правовые элементы и платформенные обязанности, не сводя их к одной процедуре.

При международной работе ARGA целесообразно собирать минимальный пакет для платформы: ссылку на конкретный материал, описание идентифицируемого лица и отсутствия согласия, категорию изображения, дату обнаружения, контакт для ответа и подтверждение полномочий представителя. Избыточное повторное направление самого интимного изображения увеличивает риск распространения и не должно быть стандартной практикой.

14. Правовые режимы: сравнительный итог

Сравнение показывает три разных регуляторных центра. Европейский AI Act прежде всего устанавливает требования прозрачности для определённых поставщиков и операторов. Французское уголовное право применяет конкретные составы к распространению манипулированного и сексуализированного материала; британское право отдельно предусматривает ответственность за создание определённых имитаций. Американский TAKE IT DOWN Act даёт специальный механизм удаления охватываемого интимного содержания на определённых платформах. Невозможно вывести из одного режима общую норму об обязательном удалении любого дипфейка.

Правильная матрица включает для каждого государства не только название закона и санкцию, но и дату действия, охраняемый объект, требования к умыслу или согласию, условия публичности, адресата обязанности, механизм немедленной защиты и доступный путь пересмотра. Для трансграничного дела отдельно фиксируются место пострадавшего, место доступности контента, сервис размещения и надлежащая юрисдикция. Эти связи значимы, но сами по себе не устанавливают подсудность либо применимое право.

15. Персональные данные, голос и идентифицируемость человека

Наличие лица, голоса или узнаваемой биографической детали в синтетическом материале не является самостоятельным ответом на вопрос о применении GDPR. Проверка начинается с того, можно ли идентифицировать человека непосредственно или косвенно, какие сведения обрабатываются, кто определяет цели и средства обработки и имеется ли территориальная связь с европейским правом. Идентифицируемость может сохраняться, даже если исходное имя заменено псевдонимом: сочетание внешности, географии, должности и редкого обстоятельства способно указать на конкретное лицо. [12]

Следует разделять сбор данных для создания модели, загрузку фотографии или голоса в сервис, публикацию результата, модерацию жалобы и хранение материалов для судебного спора. На каждом этапе возможны разные контролёры, обработчики, цели, сроки хранения и правовые основания. Согласие человека на интервью или публикацию исходного снимка нельзя без дополнительной оценки превращать в согласие на синтез его голоса для рекламы, компрометирующего ролика или фиктивного заявления.

Для независимого исследования целесообразна таблица обработки данных по каждому делу: источник, идентификаторы, основание доступа, цель, круг получателей, ограничения публикации, срок хранения и лицо, ответственное за решение о дальнейшей обработке. Такой регистр полезнее декларации «всё обезличено». Неполное закрытие лица, сохранённый голос или уникальные обстоятельства могут продолжать позволять идентификацию.

16. Данные особых категорий и сведения об уголовном преследовании

Синтетическая публикация может сообщать о состоянии здоровья, убеждениях, интимной жизни или предполагаемом уголовном поведении человека. При обработке соответствующих персональных данных необходимо оценивать не только общие статьи 5 и 6 GDPR, но и отдельные условия статьи 9, а для сведений о судимостях и правонарушениях — статью 10. Тот факт, что обвинение вымышлено, не исключает правового риска: ложное высказывание также способно обрабатывать сведения, относящиеся к идентифицируемому лицу. [12]

В материалах об INTERPOL и санкциях особенно опасно смешение источников. Фальшивая «запись допроса», синтезированное признание и скриншот несуществующего распоряжения создают видимость институционального происхождения. До любого использования необходима проверка официального первоисточника, реквизитов и канала получения. ARGA не должна помещать спорную имитацию в публичное досье под видом установленного факта.

Предлагается режим ограниченного доступа: минимизация копий, отделение текстового описания от конфиденциального медиаматериала, учет каждого просмотра и индивидуальное решение об уничтожении или архивировании. Для передачи материалов адвокату, эксперту либо государственному органу указывается самостоятельное правовое основание, а не абстрактное «в интересах расследования».

17. Право на частную жизнь, выражение мнений и общественный интерес

Свобода художественного выражения и общественная критика защищаются, но не уничтожают права на частную жизнь, репутацию и защиту личности. Проверка должна учитывать, имеет ли изображение очевидную сатирическую форму, какого человека оно касается, касается ли публикация вопроса общественного интереса, насколько велик риск введения в заблуждение и существует ли менее вредный способ донести сообщение.

Наличие подписи «пародия» может уменьшать риск обмана, однако само по себе не нейтрализует возможное унижение или раскрытие чувствительных сведений.

Нельзя, с другой стороны, объявлять дипфейком всякий материал, неудобный публичному лицу. Если единственное доказательство подделки — заявление заинтересованной стороны, исследователь должен зафиксировать конфликт показаний и пределы имеющихся технических данных. Независимая оценка требует симметричных стандартов независимо от политического, коммерческого или процессуального положения участников.

Для публикаций Observatoire рекомендуются три самостоятельные рубрики: технический статус материала; фактическая достоверность высказываний; правовая и этическая оценка распространения. Эти рубрики могут расходиться: синтетический ролик может правдиво пересказывать событие, а подлинная запись может вводить в заблуждение из-за обрезанного контекста.

18. Авторское право, смежные права и права на исполнение

Использование голосовых и визуальных образов затрагивает не только защиту личности. Исходная фонограмма, фотография, сценическое исполнение, персонаж и монтаж могут охраняться разными правами. Лицензия на использование музыкального фрагмента не равна разрешению имитировать голос исполнителя; разрешение фотографа опубликовать портрет не означает согласия изображённого лица на коммерческую имперсонацию.

В экспертном запросе необходимо установить автора исходного произведения, обладателя исключительных прав, право на изображение и голос, условия договоров с моделью или артистом, территорию и срок лицензии, разрешённые преобразования и ограничения рекламы. Отдельно анализируется использование материалов для обучения модели и создание конкретного результата: правовая оценка одного этапа не даёт автоматического ответа по другому.

Для аудита достаточна документированная цепочка разрешений. Если подписанный договор отсутствует, нельзя уверенно утверждать нарушение только из-за невозможности получить его копию; следует указать, какие подтверждения затребованы, кто отвечает за их предоставление и что остаётся недоказанным. Аналогично надпись «AI-generated» не является лицензией на чужой голос.

19. Техническая модель угроз: от генерации до переиздания

Сценарий угрозы нужно описывать цепочкой действий. Сначала злоумышленник получает исходный голос или изображение; затем создаёт модификацию; после этого подбирает канал распространения, имитирует доверенного отправителя, убирает маркеры и размещает фрагмент в правдоподобном контексте. Вред может возникнуть на любом этапе, а техническая защита одной стадии не гарантирует остальных.

Особенно опасны аудиосообщения с просьбой о срочном переводе, поддельные видеообращения должностных лиц и «сканы» документов с искусственно изменёнными реквизитами. Однако при расследовании необходимо доказать не только сходство голоса и лица. Следует проверить путь доставки, принадлежность аккаунта, историю версии файла, платёжные инструкции и наличие независимого канала подтверждения. Иногда ключевой элемент схемы — компрометация обычного почтового ящика, а не нейросетевая технология.

Модель угроз для ARGA должна включать три уровня: ложная информация о клиенте, имперсонация представителя организации и компрометация доказательств во внешнем производстве. Для каждого уровня фиксируются доступные профилактические меры, канал реагирования и допустимый уровень публичного раскрытия.

20. Проверяемое происхождение: модель утверждений и цепочка доверия

Происхождение цифрового файла — это последовательность утверждений о создании и изменениях, подкреплённых свидетельствами. Необходимо различать запись устройства, заявление редактора, подпись публикации и последующее заверение эксперта. Криптографическая проверка может подтвердить, что определённое заявление подписано соответствующим ключом и связано с байтами файла. Она не устанавливает автоматически, что подписант был очевидцем события или не ошибался. [14] [15]

C2PA предлагает инфраструктуру манифестов, включающих утверждения, связи с содержимым и электронные подписи. Проверяющий оценивает валидность манифеста, доверие к подписанту, возможные разрывы истории, использование внешних компонентов и последствия последующей редакции. Наличие валидного манифеста может быть сильным доказательством отдельного этапа обработки, но не «печаткой истины» на изображении.

Для судебно-значимой оценки следует вести отдельный протокол криптографической и контекстной проверки. Первый отвечает на вопрос, что технически подтверждено. Второй устанавливает, откуда получена запись, кем и зачем распространена, имеются ли очевидцы и могут ли независимые данные подтвердить событие. Эти процедуры не заменяют друг друга.

21. Машинно-читаемая маркировка и её пределы

Машинно-читаемая маркировка помогает автоматизированному выявлению генерации, но не делает любой алгоритм безошибочным. Маркер может быть удалён при обрезке, конвертации, пересжатии или распространении экранной записи; иногда он отсутствует потому, что система создана до соответствующего срока либо инструмент не поддерживал механизм. Поэтому отсутствие маркера нельзя считать достаточным доказательством естественного происхождения.

Следует проверять, какие именно методы использованы: криптографический манифест, незаметный цифровой водяной знак, текстовое предупреждение, метаданные файла либо

платформа, добавляющая собственную метку. У методов разная устойчивость к преобразованиям, область применимости и возможность независимой проверки. Согласно NIST, универсального решения для всех форматов и угроз не установлено. [13]

Рекомендованный отчёт описывает конкретные преобразования, при которых маркер остаётся обнаруживаемым, и пределы экспериментальной выборки. Фраза «водяной знак невозможно удалить» без испытаний, версии алгоритма и условий угроз является недопустимо категоричной.

22. Детекторы дипфейков: чего они устанавливают и чего не могут

Классификатор синтетического медиа строит вывод из признаков изображения, аудиосигнала или статистических особенностей генерации. Его результат имеет смысл лишь в контексте версии модели, набора обучения, испытательного распределения и выбранного порога. Показатель уверенности 0,92 сам по себе не означает, что вероятность подделки равна 92 процентам: часто это некалиброванная оценка модели.

В сравнительном тесте необходимо указывать базовую долю подделок, матрицу ошибок, чувствительность и специфичность, долю ложных срабатываний, доверительные интервалы, качество изображения, язык и характеристики голосов. Для открытого интернета и специально подготовленных компрометирующих материалов распределения могут резко отличаться. Нельзя переносить результат лабораторного датасета на судебную экспертизу без валидации.

При несовпадении инструментов честный ответ может звучать «недостаточно данных». Такой статус означает необходимость дополнительного исследования, а не провал эксперта. NIST рекомендует оценивать детекторы совместно с средствами происхождения, маркировки и процедурными гарантиями. [13]

23. Базовые статистические показатели независимой проверки

Для контрольной выборки фиксируются истинно-положительные (TP), ложно-положительные (FP), истинно-отрицательные (TN) и ложно-отрицательные (FN) результаты. Чувствительность составляет TP/(TP+FN), специфичность — TN/(TN+FP), точность положительного предсказания — TP/(TP+FP). Последний показатель особенно зависит от распространённости синтетических материалов в проверяемой выборке.

Условный пример: из 1 000 материалов 50 синтетические, детектор верно находит 45, пропускает 5 и ошибочно считает синтетическими 95 из 950 подлинных. Чувствительность равна 90%, специфичность — 90%, однако среди 140 положительных срабатываний настоящих подделок только 45, то есть приблизительно 32,1%. Поэтому утверждение о «90-процентной точности» может вводить в заблуждение при принятии решения о конкретном человеке.

Проверка должна разделять обнаружение синтетичности и установление злонамеренного использования. Поддельный юмористический ролик и настоящая запись противоправного поведения могут попадать в разные категории. Последующее юридическое решение нельзя автоматически выводить из результата бинарного классификатора.

24. Аудиофорензика и имитация голоса

Подмена голоса оценивается по нескольким каналам: исходной фонограмме, речевой структуре, параметрам кодирования, временным следам монтажа и внешним обстоятельствам общения. Акустическое сходство с известным человеком не подтверждает авторство. Даже биометрическая система идентификации говорящего может ошибаться при изменённом микрофоне, шуме, заболевании, сильном эмоциональном возбуждении или намеренной имитации.

Необходимо получить исходный аудиофайл без перекодирования, установить канал поступления, данные о времени и устройстве, сравнить известные образцы только при законном доступе. Декларация «нейросеть обнаружила клонирование голоса» должна сопровождаться характеристикой метода, вероятностью ошибок и перечнем альтернативных объяснений. Если доступны только сообщения из мессенджера, неполнота источника указывается прямо.

Для мошенничества практическая мера вне зависимости от экспертной проверки — обратная связь по независимому заранее известному номеру, подтверждение финансовой операции вторым лицом и пауза перед исполнением нетипичной инструкции. Эта мера снижает риск даже в случаях, когда технологию подделки установить не удалось.

25. Видео, фотографии и монтаж подлинного материала

Видеоматериал следует анализировать на уровнях контейнера, последовательности кадров, звуковой дорожки, метаданных и описания в публикации. Аномальная мимика, освещение или движения губ могут быть полезным индикатором для дальнейшей проверки, но отсутствие аномалий не доказывает подлинность, а наличие артефактов не исключает обычного монтажа, сжатия или работы камеры.

Нужно различать полностью синтетическое видео, замену лица в подлинной записи, монтаж подлинных эпизодов и ложный контекст без каких-либо изменений изображения. Последний случай демонстрирует предел чисто технического детектора: файл способен быть подлинным, но подпись о месте, времени или участнике — заведомо ложной.

Отчёт по видео должен содержать описание исходной копии, временные позиции исследованных фрагментов, сопоставленные версии, свойства перекодирования, условия проверки и перечень выводов, которые нельзя сделать. Идентификацию человека по

внешности нельзя подменять формулой «сходство с фотографией достаточно велико» без независимой методики.

26. Ошибки происхождения: отсутствующая и ложная криптографическая подпись

Подпись может отсутствовать по объективным причинам: устройство её не создаёт, платформа удаляет метаданные, материал извлечён из скриншота или архив не поддерживает формат. В таких случаях следует говорить «криптографическое происхождение не установлено». Вывод «файл точно создан ИИ» был бы необоснованным.

И обратная ситуация не менее важна. Добросовестно подписанный изготовителем файл способен содержать заведомо постановочное изображение, неверную подпись или генерацию, раскрытую самим автором. Компрометация ключа, ошибочная настройка времени или недобросовестный подписант требуют проверки идентификации сертификата и цепочки доверия. C2PA не заменяет процедуру проверки достоверности заявления. [14] [15]

Предлагается отдельное поле протокола «на что распространяется подпись». Следует записывать конкретный объект: файл, версию, набор утверждений, известный этап обработки и доверенного участника. Более широкий вывод допустим только при наличии независимых свидетельств.

27. Цепочка хранения цифрового доказательства

Для доказательственного использования важна не только техническая находка, но и возможность показать, что именно исследовалось. Протокол начинается с законного получения файла и немедленного вычисления криптографического хеша. Далее фиксируются имя, размер, формат, дата получения, исходный адрес, лицо, выполнившее копирование, способ передачи и каждое последующее действие.

Оригинал сохраняется отдельно в режиме ограниченного доступа. Анализ проводится на рабочей копии, а обнаруженные несоответствия фиксируются без «починки» исходного файла. Скриншоты, распечатки и расшифровки считаются производными материалами; они не заменяют аудио- или видеоданные, когда оригинал доступен.

Если материал получен только как URL и быстро удалён, сохранённые страницы, сообщения платформы и отметки времени могут сохранить часть фактической картины. Однако невозможность получить оригинальные байты должна уменьшать категоричность экспертного заключения. Документирование цепочки хранения не делает копию автоматически допустимым доказательством в любой юрисдикции: процессуальную оценку проводит компетентный орган.

28. Временная шкала, устройство и внешний контекст

Время создания файла, его последнего изменения, серверная дата публикации, местное время пострадавшего и время архивирования часто расходятся. Метаданные могут быть переписаны программой, устройством с неверно выставленными часами или платформой. Поэтому временная реконструкция требует указания источника каждой отметки и часового пояса.

Сопоставление независимых событий может существенно повысить или понизить вероятность заявленной истории. При проверке предполагаемого публичного выступления используются официальная программа мероприятия, регистрация присутствия, другие записи, сообщения журналистов и объективные признаки места. Но недоступность части архивов не является доказательством отсутствия события.

Формировать «единый таймлайн» допускается лишь с указанием уровня достоверности. Если последовательность остаётся гипотетической, это обозначается явно. Такая дисциплина особенно важна в экстрадиционных, санкционных и миграционных делах, где ошибка датировки может повлиять на оценку процессуальных действий.

29. Работа с платформами и уведомления о незаконном контенте

Перед обращением к платформе необходимо определить, применяется ли DSA, специальный национальный механизм либо только правила сервиса. Для уведомления по статье 16 DSA полезны точные URL, описание причины предполагаемой незаконности, доступная контактная информация и добросовестное подтверждение заявителя. Платформа рассматривает уведомление по установленной процедуре; это не означает автоматического удаления любого спорного политического или художественного содержания. [05]

Отдельно должны работать механизмы срочного удаления определённых интимных изображений, предусмотренные национальным правом. Американский 48-часовой срок связан с охватываемыми категориями материала и платформ, а не со всеми дипфейками в мире. Для Великобритании учитываются собственные обязанности сервисов и кодексы Ofcom. [10] [11]

Внутренний регистр ARGA должен сохранять дату и способ уведомления, идентификаторы материалов, полученное подтверждение, решение платформы, мотивировку отказа, действия по апелляции и риск повторной загрузки. Пересылка исходных интимных файлов многим адресатам без необходимости противоречит принципу минимизации вреда.

30. Сохранение доказательств при срочном удалении

Возникает кажущееся противоречие между требованием быстро остановить распространение и необходимостью сохранить доказательства. Оно решается разграничением публичного доступа и защищённого хранения минимально достаточной копии уполномоченным лицом.

Удаление публикации с платформы не требует уничтожения всей информации, законно сохранённой для обращения в суд, если есть надлежащее основание и меры защиты.

Однако универсальное правило «всегда скачивать исходный файл» небезопасно, особенно при содержании с участием детей, неизвестном незаконном материале или технических ограничениях доступа. В таких случаях необходимо использовать официальные способы сообщения компетентным органам, не воспроизводить изображение в общих чатах и не пытаться самостоятельно распространять опасный контент «для фиксации». Объём допустимого сохранения оценивается индивидуально.

Протокол рекомендует три уровня фиксации: только URL и внешнее описание, экранная запись доступного контекста без распространения, защищённое получение оригинала лицом, которое вправе его иметь. Выбор уровня и мотивировка заносятся в дело до передачи материалов дальше.

31. Помощь пострадавшим и недопущение вторичной травматизации

Лицу, обнаружившему интимную или унизительную имитацию, не следует предлагать многократно пересматривать изображение, пересылать его по электронной почте всем консультантам либо публиковать опровержение вместе с исходником. Вмешательство должно ориентироваться на прекращение распространения, безопасный канал связи, документирование необходимых сведений и возможность самостоятельного выбора правового способа защиты.

Практический алгоритм различает первичный ответ в ближайшие часы, сбор ограниченного пакета доказательств, уведомления платформам и последующее восстановление репутации. В срочных ситуациях, связанных с угрозами, вымогательством или несовершеннолетними, первостепенную роль играют компетентные службы и местные безопасные каналы обращения. Аналитическая организация не должна обещать невозможное «полное удаление из интернета».

Переписка с пострадавшим не должна превращать сообщение о предполагаемой подделке в расследование его личной жизни. Личные обстоятельства, не относящиеся к обращению, не включаются в досье; для публичной публикации нужна отдельная проверка необходимости и законности идентификации.

32. Международное сотрудничество: пределы компетенции частной организации

На практике одинаковый файл может находиться на серверах нескольких стран, а пострадавший, создатель и платформа — подпадать под разные правопорядки. Международная координация начинается не с выбора «самой строгой» уголовной статьи, а с

фиксации мест возникновения действий, доступности материала, местонахождения участников и юридических средств каждого государства.

Независимая ассоциация может анализировать доказательства и готовить обращения, но не располагает полномочием требовать раскрытия закрытых сведений от платформ или правоохранительных органов вне предусмотренных законом процедур. В переписке следует отделять просьбу о добровольном предоставлении данных от официального запроса компетентного органа. Недопустимо представлять экспертную рекомендацию как международное судебное решение.

Особое значение имеет минимизация трансграничной передачи персональных данных. Прежде чем отправлять медиаматериал третьей стороне, нужно определить правовое основание, страну получателя, применимые ограничения и более безопасные альтернативы, например защищённый доступ к отдельным фрагментам вместо копирования полного архива.

33. Роль синтетических материалов в делах о международном розыске

Потенциально сфабрикованная запись может быть использована как косвенное доказательство контакта, признания, публичного заявления или угрозы. В деле об INTERPOL важно отличить вопрос технической достоверности записи от юридического вопроса корректности обработки данных. Наличие имитации не доказывает автоматически, что вся международная запись незаконна; требуется установить её значение для инициирования или поддержания конкретной процедуры.

При анализе следует запросить законно доступные материалы о происхождении спорной записи, её роли в национальном деле, времени появления, способе атрибуции и последующих опровержениях. В зависимости от процедуры доказательства подаются компетентному органу или в предусмотренный механизм проверки данных. Формулировка «подделка установлена» возможна только при достаточной совокупности независимых сведений.

Рекомендация ARGA — готовить две взаимосвязанные, но разные части досье: техническую, посвящённую ограничениям и признакам обработки файла, и юридическую, посвящённую значимости этого материала для конкретной обработки данных и прав заявителя. Такой подход снижает риск смешения методов экспертизы и полномочий международной организации.

34. Санкционные совпадения, AML и ложные корпоративные заявления

Синтезированное видео руководителя компании, голосовое распоряжение о переводе или фиктивное признание связи с санкционным лицом способно попасть в коммерческую проверку контрагента. Финансовая организация, обнаружившая такой материал, должна

оценить происхождение сведений, качество внешнего источника и связь с проверяемой юридической обязанностью, а не автоматически считать публикацию доказательством.

Для AML-процедур характерна асимметрия рисков: поспешное доверие подделке может привести к необоснованной блокировке или отказу в обслуживании, но поспешное объявление настоящей записи «дипфейком» может скрыть реальный риск. Правильное решение требует подтверждения по независимым каналам, документирования причин принятого решения и предусмотренной процедуры пересмотра.

Предлагаемая форма включает идентификатор спорного материала, внешнюю ссылку, событие, конкретный риск, подтверждения из официальных реестров, результат проверки происхождения, неизвестные факты и назначенного ответственного. В отчёт не включается заведомо недоступная информация под видом «подтверждённой службой безопасности».

35. Миграция, убежище и ложные визуальные доказательства

Фальсифицированные изображения способны как необоснованно дискредитировать заявителя, так и быть использованы им самим для подтверждения преследования. Экспертиза не должна исходить из презумпции обмана одной из сторон. Необходимо учитывать отсутствие оригинальных файлов, низкое качество связи, ограничения устройств и невозможность безопасного получения дополнительных образцов.

Запрашивая дополнительные материалы, орган либо консультант должен оценить риск для заявителя и третьих лиц. Требование предоставить доступ к личным аккаунтам, устройствам или приватным изображениям не может автоматически оправдываться сомнением в одной фотографии. Альтернативными источниками могут быть официальные документы, общедоступные записи событий и подтверждения от независимых свидетелей.

В заключении недопустимы формулы «изображение не подтверждено технически, следовательно события не было». Технический вывод и общая оценка достоверности показаний остаются разными задачами. Отчёт должен указать, какой именно факт запись призвана подтвердить и какой недостаток доказательства выявлен.

36. Поддельные документы и визуальные имитации официальной переписки

Некоторые «дипфейки» представляют собой не видео с лицом, а синтезированные бланки, скриншоты писем, решения регуляторов и изображения печатей. Файл может не содержать традиционных признаков face swap, но создавать ложное впечатление официального распоряжения. Поэтому методика должна охватывать документы, интерфейсы и синтетический текст внутри изображений.

Первичная проверка устанавливает номер, дату, адресат, подпись, организационные реквизиты и соответствие форме документа. Далее ищется независимый канал

подтверждения: официальный реестр, уполномоченный отправитель, защищённый портал или заверенная копия. Нельзя делать вывод по одним орфографическим ошибкам: подлинные документы тоже могут содержать опечатки, а подделка может воспроизводить реквизиты точно.

Если документ связан с процессуальным сроком, следует отдельно сохранить первичный файл и маршрут поступления, не ограничиваясь пересланным скриншотом. Для срочной правовой реакции достаточно обозначить документ как «неподтверждённый» и запросить проверку; обвинение конкретного лица в фальсификации требует иных доказательств.

37. Стандарт языка экспертных выводов и уровни уверенности

Для независимого доклада полезна единая шкала формулировок. Наиболее определённый уровень — «подтверждено указанным первичным источником», когда проверяемое утверждение непосредственно следует из зафиксированного материала. Далее — «с высокой степенью согласуется», если несколько независимых признаков поддерживают вывод, но сохраняются альтернативы. Затем — «возможная гипотеза», «недостаточно данных» и «не подтверждено».

Термины необходимо определять операционно. «Не подтверждено» не тождественно «опровергнуто», а «признаки редактирования» не всегда означают генерацию ИИ. Для каждого заключения указываются использованные инструменты, версии, условия анализа, число проверенных файлов и основания сомнений. Противоположная сторона должна иметь разумную возможность указать на альтернативные источники или существенные ошибки метода.

Организационно важно отделить экспертное мнение от публичной позиции ARGA по делу. Если организация оказывает платные услуги заинтересованной стороне, этот факт раскрывается в объёме, необходимом для оценки независимости исследования, без разглашения защищённых данных клиента.

38. Протокол независимой рецензии и контроль смещения

Рецензент должен получить описание вопроса и минимально достаточный набор данных, не зная, насколько это возможно, желаемого исхода для инициатора. Если полной слепой проверки добиться нельзя, конфликт интересов фиксируется, а несогласие в выводах сохраняется в письменном приложении. Нельзя переписывать исследование только ради единогласия.

Рецензия проводится по пяти направлениям: полнота учёта источников; корректность процедур сохранения; валидность технических тестов; разграничение установленных фактов и предположений; соответствие юридических выводов применимому режиму. При

обнаружении серьёзного дефекта проводится повторная проверка конкретного участка, а не необоснованное признание всего материала ложным.

В итоговом протоколе отражаются замечание, его значимость, решение автора, принятые и отклонённые изменения, причины, дата и номер версии. Такая история изменений важна, когда исследование позже цитируется в суде или публичном споре.

39. Прозрачность методики и конфиденциальность проверяемого материала

Независимость требует публикации общих критериев оценки, но не требует выкладывать интимный файл, персональные контакты и все внутренние данные досье. Следует различать воспроизводимость вычисления по разрешённому контрольному набору и бесконтрольное распространение частного доказательства. Достаточным вариантом иногда является публикация алгоритма, параметров теста и обезличенной статистики при контролируемом доступе к исходным материалам для уполномоченного эксперта.

Публикуемая методика должна объяснять пороги, ограничения детекторов и порядок исправления ошибок. Если модель или инструмент недоступны для независимого повторения, это обстоятельство влияет на доказательственную силу результата и подлежит открытому указанию.

Рекомендуется двухуровневое досье: закрытое доказательственное приложение и публичная аналитическая версия. Сведения переходят между уровнями только после отдельной оценки необходимости и правомерности публикации.

40. Управление ошибками, опровержениями и правом на возражение

Публикация экспертного сообщения способна причинить дополнительный вред, если ошибочно приписывает человеку синтез материала, участие в мошенничестве или ложное заявление. Поэтому Observatoire необходимо предусмотреть механизм исправлений, позволяющий быстро отметить спорные места, сохранить исходную версию для истории и выпустить проверяемое уточнение.

Заявитель должен знать, куда направить возражение, какие материалы допустимо предоставить, кто рассматривает запрос и как устраняется конфликт интересов. Вынужденное исправление не следует маскировать незаметной заменой файла: публичная версия должна содержать историю существенных изменений без раскрытия конфиденциальной информации.

Для срочного случая допустима временная отметка «вывод оспаривается, перепроверка проводится», когда есть новая информация, потенциально меняющая решение. Это лучше, чем бездоказательное отрицание ошибки или немедленное удаление аналитического архива.

41. Профилактика в работе редакций и публичных организаций

Редакции, пресс-службы и исследовательские организации должны внедрять простой стандарт подтверждения необычных аудио- и видеообращений. Для материалов повышенного риска требуется независимый контакт с источником, проверка официального канала, фиксация происхождения и согласование редактором, не участвовавшим в подготовке материала.

Автоматические маркировки помогают, но не отменяют человеческой проверки. Изображение может быть подписано и при этом представлено вне контекста; «живое» видео со сжатием может не иметь доступного оригинала. При публикации следует прямо указывать известные технические ограничения, а не создавать видимость стопроцентной уверенности.

План реагирования включает ответственного за приём сообщений о подделках, защищённый архив, шаблон публичного уточнения и процедуру взаимодействия с пострадавшими. При повторяющихся инцидентах полезнее пересмотреть маршрут принятия редакционного решения, чем каждый раз искать один новый детектор.

42. Комплаенс разработчиков генеративных систем

Соблюдение статьи 50 AI Act требует от поставщика соответствующих систем продумать машинно-читаемую маркировку и техническую обнаруживаемость синтетического содержания в пределах нормы. Документация должна описывать область действия, поддерживаемые форматы, исключения, результаты испытаний устойчивости и остаточные риски. Эти обязанности не нужно смешивать с отдельной обязанностью профессионального пользователя раскрывать характер конкретного дипфейка. [01] [02]

В плане проверки поставщика следует фиксировать версию продукта и дату вывода на рынок. Для части систем, размещённых до 2 августа 2026 года, переходный срок исполнения обязанностей статьи 50(2) продлён до 2 декабря 2026 года; он не является общей отсрочкой всех положений статьи 50. [04]

Наличие добровольного кодекса и сертификата может быть полезным элементом комплаенса, но не освобождает от проверки фактического выполнения юридически применимых обязанностей. Условия контракта могут требовать журналирования изменений политики маркировки, уведомления о критических дефектах и отдельного порядка взаимодействия с пострадавшими.

43. Проверка устойчивости к повторному распространению

Публикация синтетического материала редко ограничивается одним файлом. Пользователи обрезают фрагмент, накладывают титры, меняют скорость воспроизведения, снимают экран телефоном или загружают ролик в другой формат. Это означает, что проверять обнаружение

и маркировку следует на типичных производных копиях, а не только на оригинале высокого качества.

Испытательный набор строится из заранее разрешённых контрольных файлов и заранее описанных преобразований. Для каждого фиксируются исходный хеш, операция, настройки кодирования, новая контрольная сумма, исходный маркер и результат проверки. Нельзя использовать конфиденциальные материалы пострадавшего в открытой демонстрации без законного основания.

Результаты лучше показывать по классам преобразований с абсолютными числами: сколько из ста тестов выдержал метод при обрезке, пересжатии, шуме и экранной записи. Средний процент без разбивки может скрыть критическую слабость именно на распространённой платформе.

44. Сравнение детекторов и контрольных наборов

Для честного сравнения алгоритмов требуется одинаковая, заранее замороженная выборка. Нельзя сначала просмотреть результаты, а затем заменить трудные файлы более удобными и публиковать новый процент как объективную оценку. Необходимо разделять обучающие, настроечные и окончательные тестовые данные, исключать дубликаты и фиксировать неизбежные пересечения.

Тест должен представлять реальные условия применения: короткие видео, разные языки, низкое качество, записи с экрана, традиционный монтаж, шумные фонограммы. Отдельно учитываются группы, для которых инструмент ошибается чаще. Если размер такой группы слишком мал, корректный вывод — недостаточная оценённость, а не голословное утверждение о равной точности.

Выбор порога зависит от цены ошибки. Для автоматической блокировки публичной публикации ложное срабатывание может ограничивать свободу выражения; для сигнала о необходимости ручной проверки приемлемый порог может быть другим. Методика должна связывать каждый порог с конкретным использованием и процедурой апелляции.

45. Риск ложного отрицания и ложного обвинения

Две противоположные ошибки одинаково разрушительны для институционального доверия. Поддельный материал может быть принят за подлинный, а подлинная запись может быть отвергнута как нейросетевая. Во втором случае возникает так называемый эффект отрицания подлинных свидетельств: сам факт существования генеративных инструментов используется как универсальная отговорка.

Чтобы избежать этого, организация не должна объявлять «подделкой» материал только из-за его политического неудобства, подозрительной эстетики или отсутствия метаданных. Надёжная оценка соединяет техническую экспертизу, происхождение файла, внешние

свидетельства и анализ контекста. Необходимо показывать альтернативные объяснения, особенно при заключениях с потенциально серьёзными последствиями.

Внутреннее решение об ограничении распространения спорного файла может приниматься при более низком уровне уверенности, чем публичное обвинение конкретного автора в фабрикации. Различные стандарты реакции должны быть заранее прописаны и не подменяться единым расплывчатым порогом.

46. Институциональная модель Observatoire ARGA

Предлагается разграничить методологическую группу, группу первичного реагирования и редакционный совет. Первая отвечает за технические протоколы и проверяемость инструментов; вторая — за законную фиксацию жалоб и минимизацию вреда; третья — за публикационные решения, конфликты интересов и исправление ошибок. Эти функции могут выполнять небольшие команды, но решения с существенными последствиями требуют обозначенного ответственного и независимого второго взгляда.

Отдельный вопрос — отношение исследовательского центра к платной клиентской практике ARGA. При заключении по делу клиента необходимо раскрывать соответствующий характер поручения, различать адвокатскую правовую позицию и независимый научный вывод, а публичный отчёт выпускать только после проверки допустимости раскрытия обстоятельств дела.

Observatoire не заявляет о праве самостоятельно накладывать санкции, удалять данные из международных систем, принуждать платформы или признавать материалы допустимыми в суде. Ценность модели заключается в прозрачной процедуре, квалификации выводов и возможности их критической проверки.

47. Этапы внедрения и критерии качества

Первый этап — подготовка процедуры приёма сообщений, шаблона согласия и правового основания обработки, матрицы компетенций и защищённого архива. Второй — ограниченный пилот на добровольно предоставленных или публичных и законно обработанных материалах, причём самые чувствительные виды контента не используются без специальных гарантий. Третий — внешняя методическая рецензия и публикация агрегированных результатов.

Критерии качества должны измерять не объём собранных изображений, а число дел с полным протоколом происхождения, долю своевременных решений, время исправления ошибок, полноту раскрытия конфликтов интересов и устойчивость средств обнаружения на независимом контрольном наборе. Любой рост числа жалоб может означать не рост вреда, а улучшение выявления: интерпретация требует осторожности.

Запуск допускается лишь при назначенном ответственном за обработку данных, описанной правовой схеме, обученном персонале и понятном порядке отказа в рассмотрении материалов, которые не могут безопасно храниться. Эти требования задаются как предлагаемый внутренний стандарт, а не как утверждение о уже проведённой сертификации ARGA.

48. Ограничения исследования и направления дальнейшей работы

Доклад представляет нормативное и методическое исследование, а не результаты самостоятельного лабораторного испытания конкретных детекторов. Приведённые числовые примеры являются условными и предназначены для объяснения статистических ошибок. Утверждения о фактической эффективности коммерческих инструментов без независимых тестов не делаются.

Правовые положения оценены по состоянию на 23 октября 2026 года. В конкретном деле необходима дополнительная проверка актуальной редакции нормы, применимых исключений, территории и даты действия. Особенно быстро меняются способы технической маркировки, доступность генеративных сервисов и практика платформ по удалению контента.

Продолжение работы возможно по трём направлениям: сравнительное тестирование алгоритмов на разрешённом наборе, независимая оценка отказоустойчивости маркировки и исследование процессуального обращения с синтетическими доказательствами в разных государствах. Для каждого требуется отдельный протокол и контроль заинтересованности исследовательской группы.

49. Обманчиво подлинная запись и стратегия отрицания реальных событий

Распространение синтетических медиа создаёт второй вид общественного риска: возможность без доказательств отвергать действительные записи, заявляя, что «сейчас всё можно сгенерировать». Эта логика особенно опасна в ситуациях, когда цифровое видео подтверждает насилие, угрозу или неправомерные действия должностного лица. Исследователь не вправе считать сомнение разумным только потому, что существуют инструменты имитации. Возражение должно иметь проверяемое основание, а не ссылаться на абстрактную технологическую возможность.

Рабочая процедура начинается с точного формулирования оспариваемого утверждения. Одно дело — автор отрицает конкретную фразу, другое — оспаривается место съёмки, третье — утверждается замена лица в нескольких секундах. Для каждой гипотезы применяются разные тесты. Полнота исходной записи, независимые копии, свидетельства очевидцев, данные о месте и времени, сведения о публикации и технические признаки обработки оцениваются совокупно. При этом отсутствие одного элемента не обнуляет автоматически остальные.

Обязательна симметрия стандартов. Если сомнения в материале одной стороны исследуются с применением строгих правил, сходный материал другой стороны оценивается по той же методике. В институциональном отчёте нельзя обозначать любую неблагоприятную для клиента запись «вероятным дипфейком» без фактических признаков и доступного протокола. Это разрушает как доверие к экспертной работе, так и возможность добросовестно представлять материалы органу правосудия.

50. Информационная безопасность при работе с экспертными инструментами

Загрузка подозрительного файла в сторонний онлайн-детектор может повлечь новую обработку персональных данных, коммерческой тайны или материалов уголовного производства. До использования сервиса следует проверить условия хранения, возможность повторного обучения моделей на загрузках, место обработки, список субподрядчиков, меры безопасности и порядок удаления. Если эти сведения отсутствуют, предпочтителен локальный инструмент либо анализ, который не требует передачи полного объекта.

Исследователь должен отличать автономный просмотр файла от исполнения активного содержимого. Вредоносная ссылка, вложение с исполняемыми элементами или специально подготовленный мультимедийный контейнер могут воздействовать на рабочую станцию. Поэтому работа проводится в контролируемой среде с ограниченными правами доступа, журналированием и раздельным хранением ключей. Риск цифровой компрометации может существовать независимо от того, является ли сам ролик синтетическим.

Протокол выбора инструмента включает название продукта, версию, правовую оценку загрузки, цели обработки, доступность удаления, стоимость и предполагаемые сроки ответа. Для особо чувствительных материалов решение принимается ответственным специалистом заранее, а не после случайного помещения файла в общедоступный сервис. Вывод инструмента хранится как отдельный результат с собственным временем и параметрами.

51. Политическая коммуникация, выборы и публичные заявления

Синтетическое обращение кандидата, министра или государственного органа может ввести аудиторию в заблуждение в течение часов, тогда как независимая техническая проверка занимает существенно больше времени. Это создает необходимость двухскоростной реакции: сначала ясно обозначить, что подлинность пока не подтверждена, затем провести документированный анализ и обновить оценку. Такая коммуникация не должна заранее обвинять политического оппонента в изготовлении материала.

Особую проблему создаёт удаление контекста при републикации. Отрывок сатирического шоу или реконструкция исторической речи могут быть распространены с подписью «запись скрытой камеры» или «официальное заявление». При анализе важно сохранить обе версии публикации, сравнить аудиторию, время и наличие предупреждения. Юридическая оценка

первоначального автора и вторичного распространителя может различаться в зависимости от роли, знания и намерения каждого.

Observatoire полезно разработать форму краткой публичной справки со следующими элементами: объект проверки; что достоверно установлено; что пока неизвестно; проверенные первоисточники; ограничения метода; время следующей редакции. Для заявления о публичном лице достаточно сведений, нужных для понимания риска, без публикации спорного файла в максимальном разрешении.

52. Недобросовестное применение маркировки и злоупотребление статусом эксперта

Технические метки могут использоваться недобросовестно. Изготовитель способен самостоятельно подписать заведомо ложную постановочную запись, а распространитель — добавить поверх подлинного видео надпись «AI-generated», чтобы аудитория перестала воспринимать его как доказательство. Следовательно, наличие и отсутствие маркировки являются сведениями о процессе публикации, но не исчерпывающим тестом истинности содержания. [14] [15]

Экспертные продукты также создают риск ложной авторитетности: опубликованная диаграмма с точным процентом может маскировать ненадёжный набор испытаний или отсутствие внешней проверки. Необходимо проверять финансирование оценки, состав тестовых объектов, версию модели, основания выбора порога и изменения настроек после получения результатов. Независимый исследователь не должен выдавать сертификат производителя за собственную экспертизу.

В публичных докладах следует отделять результаты испытания, проведённого самой организацией, от сведений, полученных от разработчика или опубликованных в исследовательской литературе. Когда лабораторных испытаний нет, необходимо прямо указать, что рекомендации имеют методический характер. Выражение «подтверждено нашими тестами» в таком случае недопустимо.

53. Экспертные ошибки при использовании визуальных эвристик

В обучающих материалах часто рекомендуют искать необычные пальцы, зеркальные отражения, нарушение синхронизации губ или ошибки освещения. Эти признаки могут быть отправной точкой, но их доказательная сила быстро уменьшается по мере развития генеративных моделей. Одновременно обычное видео способно содержать те же артефакты в результате ночной съёмки, сильного сжатия, интерполяции кадров, цифровой стабилизации или повреждения файла.

Проверяющий должен выдвигать альтернативные объяснения каждого найденного признака. Неправильное число пальцев может появиться и на обычной фотографии вследствие

перекрытия, перспективы или обработки художником. Несовпадение губ со звуком может быть вызвано задержкой трансляции. Метаданные, показывающие программу монтажа, не означают преступную фабрикацию: профессиональная редактура присутствует в обычном телевизионном производстве.

Для качественного заключения необходимо привести не просто перечень «странностей», а связь каждого признака с конкретной проверяемой гипотезой. Если инструмент исследует только визуальную часть, вывод не должен распространяться на реальность речи или события. В сомнительном случае целесообразна ограниченная формула о невозможности подтвердить происхождение.

54. Пределы применения вероятностного экспертного вывода в суде

Технический процент модели не совпадает с правовой степенью доказанности. Даже корректно калиброванный алгоритм оценивает определённый класс данных при известных условиях, тогда как суд рассматривает совокупность обстоятельств конкретного дела, допустимость доказательств и состязательные возражения. Нельзя механически переводить «0,9» в заключение «ответчик на 90 процентов изготовил подделку».

Исследовательский отчёт должен указывать, какое предположение проверялось и какая гипотеза сравнения была использована. Если детектор классифицирует синтетичность видео, он не устанавливает автора, мотив, согласие изображённого лица, способ распространения и вред. Невозможность доказать автора не отменяет результатов технического исследования, но ограничивает объём правового заключения.

При подготовке материала к представлению в производство полезно приложить краткую справку об инструменте, контрольных испытаниях, исходном материале, процедуре получения, экспертных ограничениях и зависимости результатов от качества данных. При наличии независимого рецензента его замечания следует сохранять, включая несогласие. Необходимость дополнительных процессуальных форм определяется местным правом, а не самим названием «форензика».

55. Параметры безопасной публичной демонстрации

Публикация технического отчёта может потребовать примеров дипфейков, но демонстрация реальных компрометирующих изображений способна многократно распространить вред. Для обучающих мероприятий предпочтительны специально созданные демонстрационные материалы с согласием участников, вымышленные персонажи или разрешённые тестовые наборы. Они должны быть ясно обозначены как иллюстрации, чтобы в будущем не попасть в новостную ленту без контекста.

Методическая демонстрация должна объяснять как возможности, так и ограничения технологии. Показывая удачный пример обнаружения, необходимо предъявлять и случаи

ложного срабатывания, а не формировать впечатление абсолютной надёжности. Для публичных выступлений желательно демонстрировать только обрезанные или синтетически созданные обучающие фрагменты, не содержащие настоящих личных данных потерпевших.

Если материалы используются при международных консультациях, разница между закрытым доказательственным приложением и публичным слайдом заранее согласуется. Факт наличия файла в архиве клиента не предоставляет автоматического права показывать его на конференции или загружать в программу трансляции.

56. Контроль качества переводов и расшифровок

Международный спор о дипфейке часто сопровождается переводом речи или текста. Ошибка перевода может изменить смысл утверждаемого признания, угрозы или публичного распоряжения независимо от технической подлинности записи. Поэтому перевод и экспертиза синтетичности должны проводиться по отдельным протоколам, с сохранением оригинального языка и отметками времени.

Автоматическая транскрипция может выдумывать слова в шумных фрагментах. Результат распознавания речи следует сопоставлять с аудио, отмечать неразборчивые места, различать говорящих и сохранять редакцию исправлений. Выделение красным цветом «признания» внутри машинной расшифровки без проверки профессиональным переводчиком не является достаточным подтверждением содержания.

При переводе на французский или английский нужно сохранить различие между «якобы», «по утверждению», «не подтверждено» и «опровергнуто». Эти слова имеют доказательственное значение. Одновременно указывается, кто выполнил перевод, использовалась ли автоматическая система и какая версия записи служила исходником.

57. Институциональные показатели и недопустимость ложной статистики

Количество сообщений о подозрительном контенте не равно количеству подтверждённых дипфейков. Рост жалоб может отражать улучшение работы канала обращений, медийную кампанию или изменение доверия к организации. Статистика должна различать полученные сообщения, дела, принятые к проверке, завершённые технические заключения, уточнённые и отозванные выводы, а также случаи, где сведения оказались недостаточны.

Внутренние показатели должны включать время реакции на потенциально опасный материал, процент обращений с достаточным подтверждением источника, долю обнаруженных процессуальных ошибок и число своевременных исправлений. Нельзя представлять рост процента удалений как доказательство успешного уголовного преследования: разные стадии имеют разный предмет и результат.

При публикации агрегированной статистики необходимо оценивать риск косвенной идентификации, особенно для редких дел с международным розыском или особыми

категориями данных. Малочисленные группы могут потребовать объединения категорий или отказа от публикации числа. Сводная таблица должна содержать определения и знаменатели для всех процентов.

58. Финансирование независимого аудита и контроль качества контрактов

Финансовая модель влияет на доверие к экспертному заключению. Если гонорар зависит от вывода «материал подделан», возникает прямая угроза независимости. Это не означает, что любое платное исследование бесполезно, но требует договора с заранее определённым предметом, фиксированной методикой, прозрачной стоимостью и правом эксперта сообщить результат, неблагоприятный для заказчика.

Отдельные условия регулируют доступ к исходным данным, права на методический отчёт, публикацию выводов, уничтожение рабочих копий и возможность независимой рецензии. В контракте не следует обещать доказательство факта, который ещё не исследован, или признание отчёта компетентным органом. Внимание уделяется процессуальным ограничениям конфиденциальной переписки и поручений.

При совместном участии ARGA в консультационной защите и публичном Observatoire следует разграничить право представлять позицию клиента и статус независимого исследовательского вывода. Заявления о полной организационной независимости должны опираться на проверяемую систему управления, а не на название подразделения.

59. Реестр открытых вопросов для международной консультации

Завершая исследование, целесообразно сформулировать вопросы, по которым международная практика пока не даёт исчерпывающего ответа. Как обеспечивать сохранность машинной маркировки при массовом пересжатии видеоплатформами? Как избежать превращения обязательной маркировки в косвенный запрет анонимного творческого выражения? Каким образом публичный оператор может доказать выполнение требования раскрытия, если его предупреждение систематически удаляют републикаторы?

Не менее важны доказательственные вопросы. Какие результаты тестов детекторов допустимо сравнивать между странами; как обозначать неопределённость для суда; каким образом обеспечить доступ противоположной стороны к методике без передачи интимных изображений третьим лицам? Должен ли подписанный манифест удостоверять только происхождение или также наличие редакционного контроля? Эти вопросы нужно разделить по компетенции разработчиков стандарта, государственных регуляторов и процессуальных законодателей.

ARGA Observatory может предлагать методические решения, поддерживать публичные консультации и собирать обезличенные примеры ошибок. Но публикация рекомендаций не

должна выдавать желаемые будущие нормы за уже действующие. Подобное разграничение сохраняет полезность исследования для регуляторов и защищает доверие читателя.

60. Заключение

Проблема deepfakes не сводится к вопросу, умеет ли алгоритм обнаруживать аномалии лица. Она затрагивает содержание правового высказывания, происхождение файла, доверие к цифровым каналам и возможность человека быстро защититься от ложной публикации. Технологическая маркировка необходима, но сама по себе недостаточна; удаление вредного материала необходимо в предусмотренных случаях, но не должно уничтожать законно сохраняемые доказательства; независимая экспертиза полезна, но не заменяет суд и регулятора.

Для ARGA Observatory наиболее оправдана процедурная модель: минимизировать вред, сохранить только необходимые доказательства, проверить происхождение несколькими независимыми путями, честно указать пределы технических методов и обеспечить право на возражение. Чёткое разделение установленных фактов, экспертных оценок и юридических предположений будет главным условием доверия к будущим исследованиям.

Рекомендуется утвердить стандартную матрицу квалификации контента, единый протокол фиксации и хранения, форму обращения к платформе, шкалу доказательственной уверенности и порядок рецензирования. Практические образцы приводятся в приложениях к настоящему докладу и требуют адаптации к стране, делу и компетенции исполнителя.

Приложение 1. Матрица квалификации синтетического материала

Форма предназначена для предварительного разбора, а не для установления состава правонарушения. Исполнитель заполняет признаки независимо друг от друга; пустые клетки обозначаются «неизвестно», а не «нет». Матрица особенно полезна при неполных сведениях об исходном файле, поскольку предотвращает подмену вопроса о вреде вопросом о происхождении.

Поле	Значение и правило заполнения
Объект проверки	URL, файл, архивный идентификатор, дата и способ получения
Природа преобразования	Полная генерация, замена лица, синтез голоса, обычный монтаж, не установлено
Изображённые субъекты	Идентифицируемые лица и основание такой идентификации; другие объекты
Контекст публикации	Официальное сообщение, коммерция, сатира, частный чат, предполагаемое доказательство
Уведомление аудитории	Тип маркировки, место в интерфейсе, читаемость, сохранность в копиях
Вред и предполагаемая норма	Частная жизнь, мошенничество, безопасность, авторское право, риск ошибочного решения

Поле	Значение и правило заполнения
Уровень вывода	Факт, мнение эксперта, предположение, данные отсутствуют

Завершение формы требует подписи лица, выполнившего первичную квалификацию, и отдельной отметки о необходимости передачи дела профильному юристу. Если публикация содержит интимный образ или данные ребёнка, применяется повышенный режим доступа независимо от технической оценки.

Приложение 2. Разделение ролей в соответствии со статьёй 50 AI Act

Это контрольный лист для установления адресата требований. Название договора не определяет автоматически юридическую роль: организация может одновременно предоставлять инструмент, использовать его и распространять результат. При совмещении ролей обязанности оцениваются раздельно.

Функция	Юридически значимые вопросы
Provider	Кто выводит генеративную систему на рынок; поддерживается ли машинная маркировка и обнаружение
Deployer	Кто использует систему под собственной ответственностью; используется ли она в профессиональной деятельности
Редактор/издатель	Кто принимает решение о публикации, подписи и предупреждении об искусственном содержании
Платформа	Какие посреднические функции выполняет сервис; действует ли DSA или иной специальный режим
Частное лицо	Исключена ли чисто личная непрофессиональная деятельность из соответствующей обязанности AI Act
Получатель сообщения	Какое предупреждение реально видит разумно внимательный адресат и когда он его видит

Необходимо отдельно проверить исключения для художественных произведений, уголовно-процессуального контекста и технического редактирования. Принадлежность к исключению не означает, что не применяются GDPR, уголовное право или права на изображение.

Приложение 3. Матрица дат и действия нормативных требований

Матрица фиксирует положение на 23 октября 2026 года; перед использованием в споре даты и территориальная применимость проверяются повторно. В колонке «смысл» специально разграничены начало действия требований, переходный срок и будущая обязанность государств.

Дата	Режим и правовое значение
2 августа 2026	Применение обязанностей прозрачности статьи 50 AI Act в установленных пределах

Дата	Режим и правовое значение
2 декабря 2026	Переходная дата для статьи 50(2) по определённым системам, уже размещённым на рынке; не общая отсрочка
14 июня 2027	Предельный срок транспозиции Директивы (ЕС) 2024/1385 для государств-членов
23 мая 2024	Действие редакций статей 226-8 и 226-8-1 французского УК после закона 2024 года
6 февраля 2026	Начало действия указанных CPS составов об изготовлении или заказе создания мнимого интимного изображения в Англии и Уэльсе
29 июня 2026	Начало действия указанного CPS состава об изготовлении или предоставлении специальных генераторов
19 мая 2026	Начало исполнения платформенных обязанностей раздела 3 TAKE IT DOWN Act
30 сентября 2026	Вступление описанных Ofcom мер по сопоставлению хешей для соответствующих сервисов

В докладе не допускается обобщённая фраза «на всей территории ЕС все дипфейки запрещены с августа 2026 года». Такое утверждение смешивало бы требования маркировки, будущую транспозицию специальной директивы и национальные составы ответственности.

Приложение 4. Карточка первичного сообщения об инциденте

Поступившее сообщение регистрируется без требования сразу доказать искусственное происхождение файла. Цель — определить срочность, сохранить безопасный канал и ограничить ненужную передачу чувствительных материалов. Отдельная колонка показывает, что не следует собирать без необходимости.

Обязательное поле	Практическое правило
Номер обращения	Уникальный внутренний идентификатор без полного имени в названии файла
Канал связи	Защищённый контакт, дата, язык, согласованный способ обратной связи
Обнаружение	Ссылки, время и описание угрозы; копию интимного материала не требовать по умолчанию
Срочность	Угроза физической безопасности, вымогательство, ребёнок, активный мошеннический перевод
Первичная цель	Удаление, опровержение, техническая проверка, подготовка обращения
Доступ	Кто вправе просматривать материал, кто отвечает за срок хранения
Направление	Юрист, компетентная служба, платформа, техэксперт или прекращение рассмотрения

Сотрудник сообщает, что организация может оценить материал и подготовить документы, но не гарантирует удаления всех копий. При необходимости экстренной защиты пострадавшего перенаправляют в компетентные службы, не затягивая помощь ради лабораторного анализа.

Приложение 5. Минимальный цифровой паспорт файла

Паспорт связывает анализ с конкретным объектом. Он составляется сразу после законного получения и должен позволять другому специалисту найти ровно ту версию, которая изучалась. Несовпадение хешей требует расследования причин, а не молчаливой замены объекта.

Реквизит	Что фиксировать
Обозначение	Имя исходника и внутренний идентификатор объекта
Получение	Законный источник, уполномоченное лицо, канал, дата, часовой пояс
Размер и формат	Байты, MIME, контейнер, кодеки аудио и видео при наличии
Контрольная сумма	Алгоритм SHA-256, значение и инструмент вычисления
Исходный контекст	URL, описание страницы, подпись, сведения об авторе заявления
Производные версии	Рабочие копии, преобразования, отдельные хеши
Условия доступа	Доступные роли, основание хранения, срок пересмотра

Если удалось получить лишь скриншот, в паспорте он описывается именно как скриншот: это не подменяет оригинальную запись. В публичный отчёт полный хеш можно включать только при оценке риска сопоставления с чувствительным материалом.

Приложение 6. Реестр цепочки хранения

Каждое действие с оригиналом или его копией должно оставлять запись. Это позволяет объяснить, были ли изменения результатом законной обработки, обычной перекодировки или возможного вмешательства. Ниже приведён минимальный реестр; дополнительные поля вводятся требованиями конкретного производства.

Действие	Доказательная запись
Приём	От кого, на каком основании и каким техническим каналом
Контроль целостности	Проверка хеша при получении и при передаче
Копирование	Исполнитель, цель копии, местонахождение оригинала
Анализ	Инструмент, версия, параметры, полученный результат
Передача эксперту	Основание, защищённый способ, квитанция о получении
Публикация	Разрешённая редакция, исключённые поля, лицо, одобрившее раскрытие
Архивирование/уничтожение	Основание срока, дата, метод и подтверждение ответственного

Изменения в журнале исправляются дополнительной строкой, сохраняя первоначальную запись. Сам журнал должен храниться отдельно от общедоступного отчёта.

Приложение 7. Протокол проверки C2PA и других сведений о происхождении

Протокол требует установить не только наличие манифеста, но и результат криптографической проверки. Если инструмент показывает предупреждение, оно переносится в отчёт дословно в пределах допустимого цитирования, а оценщик объясняет его значение. Проверка проводится на неизменённой копии.

Шаг	Контрольный вопрос
Обнаружение	Встроен ли манифест и доступна ли связанная внешняя информация
Подпись	Успешно ли проверяется цифровая подпись данного манифеста
Цепочка доверия	Известен ли подписант и на каком основании ему доверяют
Связь с файлом	Совпадают ли заявленные привязки с данным объектом
История	Какие изменения и исходные компоненты подтверждены, а где имеются пробелы
Контекст	Можно ли независимо подтвердить описанное событие, автора заявления и время
Итог	«Подтверждён манифест» отдельно от «подтверждено событие»

При отсутствии C2PA ставится отметка «история происхождения данным способом не установлена». Запрещается превращать её в заключение о генерации.

Приложение 8. Паспорт проверки машинной маркировки

Проверка должна охватывать фактический рабочий формат, а не рекламную демонстрацию производителя. Исполнитель заранее определяет набор преобразований и допустимый процент потерь обнаружения. При работе с коммерческим инструментом требуется сохранить условия лицензии и сведения об ограничениях доступа.

Область	Проверяемая характеристика
Исходный маркер	Тип, версия, носитель, место размещения, способ чтения
Аудио/видео	Поддерживаемые контейнеры, кодеки, частота дискретизации
Устойчивость	Сжатие, монтаж, обрезка, наложение текста, шум, запись с экрана
Ложные срабатывания	Результаты на подлинном контрольном наборе
Ложные пропуски	Результаты на разрешённом синтетическом наборе
Обход	Какие угрозы проверялись и какие не проверялись

Область	Проверяемая характеристика
Вывод	Граница между техническим испытанием и юридической обязанностью

Предпочтительно публиковать отдельно исходные количества успешных и неуспешных испытаний и метод отбора. «Обнаруживается в 97% случаев» без описания состава случаев не является полноценной характеристикой.

Приложение 9. Бланк независимого теста детектора

Испытание начинается с письменной постановки задачи. Одно дело — отбор подозрительных файлов для ручной проверки, другое — блокировка публикации или формирование экспертного заключения. Требования к ошибкам здесь различаются.

Поле	Запись
Версия метода	Название, поставщик, версия, дата, контроль настроек
Набор данных	Законное происхождение, количество, классы и ограничения
Исключения	Дубликаты, близкие версии, контент вне предмета теста
Порог	Значение, выбранное до финального теста, и обоснование
Результат	TP, FP, TN, FN с абсолютными числами и долями
Неопределённость	Доверительный интервал, редкие подгруппы, переносимость
Ограничение использования	Сценарии, в которых результат нельзя использовать самостоятельно

Для сравнения двух моделей следует заранее определить критерий победы и запретить изменение тестового набора после знакомства с оценками. Приоритет отдается воспроизводимому эксперименту, а не числу знаков после запятой в рекламной таблице.

Приложение 10. Проверка статистики на условном примере

В учебной выборке из 1 000 единиц 50 являются синтетическими. Детектор верно определяет 45, пропускает 5, ошибочно маркирует 95 подлинных и верно принимает 855 подлинных. Все числа условны и не являются результатом испытания оборудования ARGA.

Метрика	Расчёт и результат
Чувствительность	45/(45+5) = 90%
Специфичность	855/(855+95) = 90%
Точность положительного прогноза	45/(45+95) ≈ 32,1%
Точность отрицательного прогноза	855/(855+5) ≈ 99,4%
Общая доля верных решений	(45+855)/1000 = 90%
Вывод	Положительный сигнал требует независимой проверки, особенно при низкой распространённости подделок

Этот пример объясняет, почему «точность 90%» не означает, что девять из десяти обвинённых материалов действительно синтетические. Для реального заключения необходимы базовые частоты проверяемой среды и валидация порога.

Приложение 11. Контрольный лист аудиозаписи

Для каждого исследуемого аудио устанавливаются тип сообщения, его происхождение и применимый канал доступа. До применения акустических средств оценивается возможность подтвердить разговор через прямой контакт сторон. Этот шаг часто полезнее сложного сравнения спектрограмм.

Проверка	Что должно быть указано в протоколе
Первичный объект	Оригинальный контейнер, кодек, длительность и хеш
Источник	Кто и когда записал, как передал, есть ли исходное устройство
Редактирование	Возможные склейки, повторные перекодирования, вставки
Сравнение голоса	Законно полученные эталоны, ограничения и риск смешения голосов
Внешняя верификация	Участники разговора, независимый контакт, журнал звонков при законном доступе
Итог	Что подтверждено, что спорно и какие дополнительные материалы нужны

Публичное раскрытие образцов голоса требует отдельной проверки необходимости и законности. Нельзя обучать на голосах участников новую модель без соответствующего основания только ради демонстрации схожести.

Приложение 12. Контрольный лист видеозаписи

Видео исследуется с сохранением последовательности кадров и звуковой дорожки. Особое внимание уделяется тому, не является ли ошибочное заключение следствием обычной стабилизации, HDR, повышения резкости или сильного сжатия платформы.

Компонент	Проверка
Контейнер	Параметры, хеш, наличие отдельной звуковой дорожки
Кадры	Частота, пропуски, повторения, признаки перекодирования
Аудиовизуальная связь	Согласование звука с артикуляцией, возможная задержка канала
Свет и геометрия	Альтернативные естественные причины артефактов
Контекст	Место, дата, автор публикации, полные и обрезанные версии
Внешний материал	Независимые записи события, программы и свидетельства

Компонент	Проверка
Вывод	Раздельно тип монтажа, источник, достоверность утверждаемого события

Кадры, выбранные в качестве примеров, снабжаются временными метками; нельзя строить всё заключение по нескольким эстетически подозрительным кадрам без контрольного анализа.

Приложение 13. Бланк уведомления по статье 16 DSA

Форма адресована хостинг-сервису, подпадающему под применимый режим. Она не подменяет специальную уголовную жалобу и не гарантирует признание материала незаконным. Юридическое основание и точные URL должны быть сформулированы применительно к конкретной публикации.

Элемент	Пример содержания, требующий адаптации
Адресат	Название сервиса и его публичный механизм notice-and-action
URL	Полный адрес спорной публикации и при необходимости точное место в сервисе
Причина	Почему конкретная информация предположительно незаконна по применимому праву
Контекст	Кто изображён, в чём отсутствие согласия, какие последствия и ссылки на подтверждения
Заявитель	Необходимые имя/адрес электронной связи с учётом применимых исключений
Добросовестность	Подтверждение действительности изложенных сведений по знанию заявителя
Просьба	Рассмотреть уведомление, сообщить решение, сохранить законно доступную запись обработки

Если заявитель не желает раскрывать адрес проживания, он не добавляется просто «для солидности». Если материал относится к особой защищённой категории, используется соответствующая процедура и минимум данных.

Приложение 14. Срочная инструкция по интимным материалам

Эта форма задаёт порядок, минимизирующий вторичное распространение изображения. При наличии несовершеннолетнего, угрозы или вымогательства применяется специальный маршрут обращения к компетентным службам; конфиденциальный файл не рассылается экспертам ради обычного согласования.

Этап	Ожидаемое действие
Контакт	Проверить безопасный канал, срочность и полномочие представителя
Фиксация	Сохранить необходимые URL и сведения без избыточного тиражирования

Этап	Ожидаемое действие
Сервис	Применить доступную форму жалобы на интимный контент
Право	Проверить применимость местного механизма удаления и его условия
Доказательства	Законно сохранить минимальный пакет уполномоченным лицом
Повторное размещение	Организовать последующую проверку ссылок и известных копий
Уведомление человека	Объяснить, что удаление с одной платформы не гарантирует прекращения всех копий

Нельзя обещать пострадавшему гарантированный результат или требовать предоставления неопубликованного интимного исходника, если задача решается без него.

Приложение 15. Сравнительный лист механизмов защиты в ЕС, Франции, США и Великобритании

Матрица не претендует на исчерпывающий обзор судебных средств каждой страны. Она показывает, что различаются регулируемое действие, обязанное лицо и сроки; выбор механизма зависит от обстоятельств и местного права.

Режим	Основной предмет
AI Act, ст. 50	Прозрачность и маркировка определённых систем и публикаций; не универсальное удаление
DSA, ст. 16	Механизм сообщения о предположительно незаконном контенте конкретному хостинг-сервису
Франция, ст. 226-8	Определённые нераскрытые монтажи и алгоритмическая имитация без согласия
Франция, ст. 226-8-1	Отдельный состав для распространения сексуализированного материала без согласия
США, TAKE IT DOWN	Специальный порядок удаления охватываемых интимных изображений по действительному запросу
Англия и Уэльс	Отдельные составы изготовления или запроса создания определённых интимных имитаций
Ofcom / OSA	Обязанности соответствующих сервисов по предотвращению распространения незаконного интимного контента

Перед любым обращением указываются дата, территория, круг лиц и доказательства фактических элементов. Выдержка из таблицы не заменяет профессиональную оценку дела.

Приложение 16. Паспорт правомерной обработки персональных данных

Защита пострадавших не даёт организации карт-бланш на хранение любых медиа. В каждом деле необходимо определить законный источник, цель, круг доступа и срок удаления. Особое внимание уделяется изображениям, раскрывающим интимную жизнь или данные, связанные с обвинениями в преступлении.

Требование	Минимальная запись
Контролёр	Кто устанавливает цель и средства обработки
Основание	Статья 6 GDPR либо применимый альтернативный режим
Дополнительное условие	При необходимости статьи 9/10 GDPR или нормы национального права
Минимизация	Почему недостаточно текста, URL или ограниченного фрагмента
Доступ	Права, журнал открытия и ограничения копирования
Срок	Событие пересмотра и политика удаления/архива
Передача	Кому, в какую страну и по какому основанию

Не следует произвольно проставлять согласие как универсальное основание, особенно когда в деле фигурируют другие изображённые люди или данные органов уголовного преследования.

Приложение 17. Контроль публикации аналитического заключения

Публичный документ должен позволять аудитории понять вывод без просмотра защищённого интимного файла. Редактор проверяет, не создаёт ли формулировка самостоятельный репутационный вред из-за чрезмерной определённости или ненужного воспроизведения обвинений.

Блок	Контроль перед публикацией
Заголовок	Не заявляет доказанной подделки, если имеется лишь вероятностный вывод
Метод	Указаны инструменты, версии, выборка и ограничения
Персональные данные	Исключены сведения, не нужные для общественного интереса
Конфликт интересов	Раскрыт характер связи с заказчиком в допустимом объёме
Право на возражение	Доступен канал и процедура исправления
Приложения	Проверена необходимость каждого изображения, скрина и адреса
Архив	Сохранён идентификатор редакции и история существенных исправлений

Если независимое заключение невозможно без раскрытия защищённого материала, предпочтительно подготовить закрытую версию для компетентного адресата и ограниченную публичную справку.

Приложение 18. Матрица конфликтов интересов эксперта

Финансовая или процессуальная заинтересованность не всегда исключает проведение исследования, но должна быть известна лицу, оценивающему его доказательную силу.

Особенно важно обозначать, оплачивает ли заключение представитель стороны, в чью пользу сформулирован вопрос.

Риск	Контрольная мера
Оплата заинтересованной стороной	Фиксация заказчика и предмета поручения; запрет оплаты за желаемый вывод
Совместное руководство	Независимый рецензент и письменная проверка выводов
Предыдущая позиция эксперта	Декларация ранее опубликованных оценок по спорному файлу
Доступ к данным только одной стороны	Отметка ограничений и перечня невыслушанных альтернатив
Коммерческие отношения с разработчиком	Раскрытие поставщика, лицензии и ограничения доступа к алгоритму
Публичная кампания	Разделение экспертной оценки и advocacy-коммуникаций

Отсутствие конфликта в форме «ничего не указано» не считается проведённой проверкой. Декларация подписывается и хранится вместе с протоколом назначения эксперта.

Приложение 19. Техническая карта риска для финансовых операций

При голосовом или видеообращении с требованием необычного платежа техническая экспертиза не должна задерживать оперативную защиту. Организация действует по заранее согласованным правилам проверки полномочий, не раскрывая предполагаемому злоумышленнику внутренние признаки контроля.

Сигнал	Действие без признания записи подделкой
Необычная срочность	Пауза и второе независимое подтверждение
Новый счёт получателя	Подтверждение через ранее известные реквизиты и контакт
Голосовое распоряжение	Проверка личности по отдельному каналу, а не по сходству голоса
Копия корпоративного документа	Сверка с официальным реестром и уполномоченным сотрудником
Подозрительная медиаатрибуция	Сохранение законно доступного оригинала и метаданных
Последующий спор	Протокол причинного пути и решение о корректировке внутреннего контроля

Вывод «попытка мошенничества доказана» требует самостоятельных фактов и правовой оценки; подозрительное аудио само по себе недостаточно.

Приложение 20. Структура доказательственного досье по международному розыску

Форма предназначена для осторожного анализа предполагаемого синтетического «признания» или иного изображения, которое могло быть включено в национальное дело. Ни

одна графа не создаёт автоматического основания для удаления данных из международной системы.

Часть досье	Содержание
Происхождение материала	Источник, дата, связь с национальным производством
Технический объект	Оригинал, хеш, производные файлы и качество
Атрибуция	Кто объявил запись подлинной, на каких основаниях
Внешние подтверждения	Документы, время, свидетельства, независимые записи
Процессуальная роль	Какие конкретные выводы орган опирал на спорный материал
Альтернативы	Неустановленные обстоятельства, возможности монтажа без ИИ
Просьба	Конкретная проверка или мера компетентного органа с указанием применимых правил

Если запрос направляется в предусмотренный механизм INTERPOL, он должен соответствовать его процедуре и предмету проверки; частное техническое мнение не подменяет решение такого органа.

Приложение 21. Протокол оценки материалов для убежища

В делах об убежище качество мобильных медиа может быть низким, а дополнительные образцы могут подвергать заявителя или свидетеля риску. Это влияет на разумность требований к проверке, но не оправдывает отказ от анализа существующих источников.

Вопрос	Принцип оценки
Источник файла	Как получен, почему оригинал может быть недоступен
Контекст	Что конкретно запись подтверждает и где её пределы
Сомнения	Названы технические альтернативы и их вероятность без категоричности
Дополнительные материалы	Возможно ли их безопасное получение без давления на третьих лиц
Заключение	Не подменять оценку всей истории проверкой одного медиафайла
Доступ	Конфиденциальность заявителя и лиц в стране происхождения

Если результатом анализа является лишь недостаточность качества, это указывается напрямую. Правдоподобие личного рассказа и технические характеристики файла должны рассматриваться раздельно.

Приложение 22. Контроль фальшивых официальных писем и скриншотов

Цифровой документ может быть синтезирован без использования лица или голоса. Для таких объектов необходимо проверять как содержимое, так и путь передачи. Важно исключать поспешную атрибуцию нарушения автору истинного бланка.

Поле	Проверка
Идентификаторы	Номер, дата, реквизиты учреждения, ссылки на процедуру
Канал	Домен, электронная подпись, официальный портал, маршрут передачи
Визуальные признаки	Шрифты и печати рассматриваются только как вспомогательные сигналы
Прямое подтверждение	Независимое обращение к компетентному учреждению
Сроки	Нельзя пропускать процессуальный срок в ожидании технического исследования
Вывод	«Подлинность не подтверждена» отдельно от «доказана фабрикация»

Оригинальный файл сохраняется до технической редакции. Перерисовывать подписи, исправлять поля или добавлять реквизиты «для читаемости» на исследуемой копии недопустимо.

Приложение 23. Рабочий протокол апелляции и исправления отчёта

Процедура должна работать независимо от того, кто заказал исследование. Заявитель вправе указать на неправильный файл, новый первоисточник, ошибку инструмента и искажение его позиции. Ответственный фиксирует входящее сообщение и степень влияния замечания на первоначальный вывод.

Стадия	Запись
Регистрация	Идентификатор жалобы, дата, связанная версия доклада
Основание	Ошибка факта, ошибки анализа, нарушение конфиденциальности, иное
Независимый просмотр	Назначенный рецензент, проверенные материалы
Временная мера	При необходимости пометка оспариваемого вывода
Решение	Оставить, уточнить, исправить либо отозвать конкретную формулировку
Публикация	Видимая история изменений без лишних персональных данных
Завершение	Мотивировка и направление ответа заявителю

При существенной ошибке правильное исправление важнее сохранения внешнего впечатления безошибочности. Удаление критики без мотивировки недопустимо.

Приложение 24. Условия экспертного поручения и границы исследования

Договор или постановка задачи не должен требовать заранее желаемого результата. При невозможности законно получить исходные данные эксперт обязан указать этот недостаток; допустимый объём заключения должен уменьшаться соответственно.

Условие	Рекомендуемая формулировка
Предмет	Исследование указанных файлов и конкретных гипотез
Метод	Применяемые инструменты, возможность изменения при обосновании
Независимость	Запрет зависимости вознаграждения от направления вывода
Данные	Законное происхождение, режим доступа и передача
Результат	Отчёт с фактами, альтернативами, ограничениями и приложениями
Рецензия	Допустимость независимой методической проверки
Публикация	Отдельное письменное решение с учётом прав третьих лиц

Не следует включать обещание, что вывод будет признан судебным доказательством, принят международной организацией или послужит обязательным основанием для удаления информации.

Приложение 25. Матрица критериев запуска пилотной программы ARGA

До пилотирования нужен не рекламный набор инструментов, а работающий минимум управления: кто вправе получать материалы, где они хранятся и как пресекается недопустимый сбор. Эффективность оценивается после небольшой контролируемой выборки и независимой проверки.

Критерий	Условие допуска
Мандат	Письменно очерчены экспертные функции, исключены властные претензии
Законность	Определены основания обработки и специальных категорий данных
Безопасность	Управление доступом, шифрование и ограничение копий
Методика	Версионированные протоколы, пороги, порядок пересмотра
Кадры	Назначенные сотрудники, обучение конфиденциальности и ограничениям инструментов
Публичность	Политика публикации и исправления ошибок
Аудит	Внешняя рецензия до публичных обобщений об эффективности

Отдельно предусматривается критерий остановки: выявленный незаконный сбор, утечка чувствительного материала либо невозможность достоверно определить происхождение контрольного набора.

Приложение 26. Дорожная карта на 90 дней

Сроки являются предлагаемым внутренним планом и не подтверждают, что соответствующие мероприятия уже выполняются. Рабочий график учитывает приоритет безопасного приёма сообщений перед масштабированием анализа.

Период	Контрольный результат
Дни 1–15	Утверждены мандат, роли, политика приёма и хранения, матрица правовых оснований
Дни 16–30	Настроен защищённый канал, журнал доказательств и журнал запросов на исправление
Дни 31–45	Испытаны процедуры контрольного сохранения, маркировки и проверки происхождения
Дни 46–60	Проведена рецензия обезличенных тренировочных кейсов
Дни 61–75	Подготовлены публичная методика, формы уведомления и сценарии реагирования
Дни 76–90	Независимая оценка рисков, решение о масштабе пилота и ограничениях

Решение об открытии нового направления должно фиксировать не только достигнутые задачи, но и оставшиеся правовые ограничения. Любое ускорение не оправдывает работу с незаконно полученными файлами.

Приложение 27. Условный кейс: голосовое распоряжение о переводе средств

Факты учебного примера: финансовый сотрудник получает аудиосообщение с голосом, похожим на руководителя; сообщение предписывает перечисление суммы новому контрагенту. Качество записи среднее, исходный аккаунт неизвестен, а вложение передано несколько раз. Это вымышленная ситуация, не описание реального клиента ARGA.

Вопрос	Корректный рабочий ответ
Нужно ли выполнять перевод	Нет, до независимого подтверждения полномочий и реквизитов
Что сохранить	Законно доступный файл, маршрут поступления, метаданные и хеш
Что проверить	Обратный звонок по известному контакту, договор, новый счёт, список полномочий
Что может установить детектор	Лишь признаки синтетической обработки с указанными ограничениями
Каков итог	Отдельный вывод о платёжном риске и отдельный вывод о происхождении аудио

Даже если технический анализ не сможет установить генерацию, внутренний контроль способен остановить платёжный риск. Именно это отличает практически полезную реакцию от попытки сперва доказать конкретную нейросетевую технологию.

Приложение 28. Условный кейс: фальшивое признание по международному делу

Факты учебного примера: в медиа появляется ролик с человеком, предположительно признающим участие в преступлении. Государственный орган ссылается на него в публичном комментарии, но копия первоисточника недоступна. Заинтересованная сторона называет запись дипфейком. Пример вымышленный; выводов о каком-либо действительном разыскиваемом лице не делается.

Вопрос	Необходимое исследование
Идентичность	На каком основании лицо в ролике отождествляется с человеком
Источник	Кто впервые опубликовал, есть ли полный ролик или носитель
Технология	Какие признаки изменений установлены, какие альтернативы возможны
Значение	Повлияла ли запись на конкретное правовое решение и как именно
Процедура	Какой орган уполномочен рассмотреть возражение и какие документы требуются

Вывод «заявление выглядит синтетическим» не равнозначен автоматическому прекращению международной процедуры. Техническая оценка должна быть встроена в правовую аргументацию по конкретным данным и доказательствам.

Приложение 29. Условный кейс: компрометирующая интимная имитация

Факты учебного примера: совершеннолетняя заявительница обнаруживает распространённое без согласия синтетическое интимное изображение, похожее на неё. Доступны несколько URL, но изначальный источник неизвестен. Ситуация вымышлена и не основана на реальном частном материале.

Вопрос	Приоритет
Безопасность	Ограничить контакт с распространителем и оценить угрозы или шантаж
Фиксация	URL и необходимый контекст без многократного пересылания изображения
Удаление	Использовать специальный механизм платформы и применимое национальное право
Законность хранения	Установить узкую цель, доступ, срок и допустимую копию
Дополнительная защита	Рассмотреть обращение в компетентные органы и сопровождение пострадавшего

Вопрос	Приоритет
Коммуникация	Не публиковать исходник в опровержении и не обещать уничтожения всех копий

Публикация публичной заметки об инциденте возможна лишь при самостоятельном решении о необходимости и конфиденциальности. Сохранение психологической и физической безопасности важнее медийной убедительности отчёта.

Приложение 30. Условный кейс: сфабрикованное официальное письмо

Факты учебного примера: организация получает изображение письма, якобы исходящего от иностранного регулятора, с требованием срочно передать клиентские документы. Визуально документ соответствует фирменному бланку, но оригинального файла и подтверждения канала нет. Пример вымышленный.

Вопрос	Действие
Поручение	Не передавать документы до надёжной проверки полномочий
Канал	Независимая сверка официального контакта и реестра
Экспертиза	Зафиксировать изображение, метаданные и путь поступления
Предположение о подделке	Не делать категоричный вывод только по изображению печати
Обращение	При необходимости уточнить статус непосредственно у компетентного органа

Сложность такого случая заключается не в наличии синтетического лица, а в ложной институциональной атрибуции. Поэтому методика анализа дипфейков должна охватывать документы и интерфейсы.

Приложение 31. Шкала доказательственных выводов

Таблица предназначена для унификации языка, а не для перевода технической вероятности в юридический стандарт без обоснования. Категории должны использоваться одинаково в публичных и закрытых версиях.

Категория	Значение
Установленный технический факт	Непосредственно подтверждён доступным и сохранённым объектом
Сильное согласование признаков	Несколько независимых признаков поддерживают гипотезу; альтернативы описаны
Ограниченная поддержка	Часть данных согласуется с гипотезой, но выборка неполна
Недостаточно данных	Вывод о происхождении не может быть сделан добросовестно
Не подтверждено	Представленного источника недостаточно для конкретного утверждения

Категория	Значение
Опровергнуто	Независимые подтверждения показывают несовместимость проверяемого утверждения с установленными фактами

Запрещается трактовать первые три категории как судебное решение об ответственности. В каждом случае необходимо назвать предмет вывода и его адресата.

Приложение 32. Реестр рисков и корректирующих мер

Реестр обновляется по итогам фактических инцидентов и проверок. Он должен помогать назначать владельца риска, а не создавать видимость строгой количественной науки из субъективных оценок. Балльные шкалы допускаются только при определённых критериях.

Риск	Мера контроля
Ложное обвинение в генерации	Независимая рецензия и обязательное описание альтернатив
Утечка конфиденциального изображения	Ограничение доступа, защита хранилища, журналирование
Потеря исходного файла	Неизменяемая копия, хеш и резервирование в рамках закона
Некорректная статистика детектора	Независимая выборка, матрица ошибок, базовая частота
Конфликт интересов	Декларация финансирования и отдельное решение редакции
Повторное распространение при опровержении	Публикация без репоста спорного медиаматериала
Подмена роли регулятора	Ясное описание мандата и ограничений компетенции

Для каждого риска в рабочем экземпляре добавляются ответственное лицо, дата пересмотра и критерий остаточной приемлемости. Публичная версия не должна раскрывать внутренние меры безопасности так, чтобы облегчать обход.

Приложение 33. Контрольный лист научной рецензии

Независимый читатель подтверждает или оспаривает не общее впечатление от доклада, а конкретные выводы и доказательства. При ограниченном доступе к исходным данным рецензент указывает, какие элементы он смог проверить и какие получил только из пересказа автора.

Проверяемое качество	Результат рецензента
Источники	Первичные нормативные тексты и точные даты исполнения
Сбор данных	Законность, источник и воспроизводимая цепочка хранения
Детекторы	Версия, валидация и пределы переноса результатов
Статистика	Правильность абсолютных чисел, долей, интервалов и выборок
Конфликт	Проверенное раскрытие заинтересованности

Проверяемое качество	Результат рецензента
Отчёт	Различие факта, технической оценки и юридического предложения
Права лиц	Механизм возражения, исправления и минимизация публикации

Замечания рецензента сохраняются вместе с мотивированными ответами автора. Несогласие относительно интерпретации может быть опубликовано, если не раскрывает защищённые данные.

Приложение 34. Перечень запрещённых методических сокращений

Некоторые формулы выглядят убедительно, но логически не выдерживают проверки. Данный лист предназначен редактору и эксперту как контроль качества перед подписанием отчёта.

Некорректная формула	Корректная замена
«Нет C2PA — значит дипфейк»	«Манифест происхождения не обнаружен»
«C2PA подтвердил событие»	«Валидирована подпись определённых заявлений о файле»
«Детектор дал 93%, значит доказано»	«Оценка модели требует калибровки и независимой валидации»
«AI Act требует удалить любой deepfake»	«Статья 50 вводит определённые обязанности прозрачности»
«48 часов на удаление во всех странах»	«Специальный срок применяется в границах конкретного закона»
«Файл подлинный, потому что в нём нет артефактов»	«Доступные признаки не позволили установить редактирование»
«Файл смонтирован, значит событие вымышлено»	«Необходимо оценить исходное событие и значение монтажа отдельно»

Список проверяется перед каждой внешней публикацией; он уменьшает риск произвольного расширения заключений на ситуации, которых технический тест не охватывал.

Приложение 35. Термины для единообразного использования

В рабочей документации рекомендуется закрепить словарь и ссылаться на него при формировании форм. Термины не подменяют юридические дефиниции соответствующей юрисдикции.

Термин	Рабочее значение
Синтетический контент	Материал, созданный или существенно преобразованный средствами генерации
Deepfake	Имитация объектов или событий, способная создать ложное впечатление подлинности
Имперсонация	Представление содержания как исходящего от определённого лица
Provenance	Проверяемые сведения о создании и изменениях объекта

Термин	Рабочее значение
C2PA manifest	Структура утверждений и подписей, связанных с конкретным объектом
Watermark	Встроенный сигнал, рассчитанный на последующее обнаружение
False positive	Подлинный объект ошибочно распознан как синтетический
False negative	Синтетический объект ошибочно отнесён к подлинным
Chain of custody	Документированная последовательность получения, хранения и передачи

При переводе на английский или французский предпочтительно сохранять идентичные идентификаторы карточек, чтобы исключить изменение смысла при оформлении международного пакета.

Источники и нормативные материалы

[01] Регламент (ЕС) 2024/1689, редакция по состоянию на 27 июля 2026 г.; статьи 3, 50, 99, 113. Официальный источник

[02] Европейская комиссия. Guidelines on transparency obligations, 20 July 2026. Официальный источник

[03] Европейская комиссия. Code of Practice on Transparency of AI-generated Content, 2026. Официальный источник

[04] Европейская комиссия. FAQ по статье 50 и переходному сроку. Официальный источник

[05] Регламент (ЕС) 2022/2065, Digital Services Act, ст. 16–17 и другие положения. Официальный источник

[06] Директива (ЕС) 2024/1385 о противодействии насилию в отношении женщин. Официальный источник

[07] Code pénal (France), article 226-8, редакция 2024 г.. Официальный источник

[08] Code pénal (France), article 226-8-1, редакция 2024 г.. Официальный источник

[09] Crown Prosecution Service. Cybercrime prosecution guidance, обновления 2026 г.. Официальный источник

[10] Ofcom. Enforcement programme on non-consensual intimate imagery, 9 September 2026. Официальный источник

[11] Federal Trade Commission. Complying With the TAKE IT DOWN Act, 2026. Официальный источник

[12] Регламент (ЕС) 2016/679 (GDPR), статьи 5, 6, 9, 10, 12, 17 и 21. Официальный источник

[13] NIST AI 100-4. Reducing Risks Posed by Synthetic Content, 2024. Официальный источник

[14] C2PA. Content Credentials Technical Specification, версия 2.2. Официальный источник

[15] C2PA. Content Credentials Explainer, версия 2.2. Официальный источник

Примечание об источниках. Нормативные положения и методические рекомендации приведены в редакциях, доступных к 23 октября 2026 года. Иллюстративные сценарии и числовые расчёты не являются полевыми исследованиями ARGA. В конкретном производстве необходима проверка применимой редакции закона, процессуальной формы и полномочий адресата.